

Bias Corrected Bootstrap Bandwidth Selection

By

Birgit Grund, Dept. of Applied Statistics, University of Minnesota
Jörg Polzehl, Konrad Zuse Center for Information Technology¹

School of Statistics, University of Minnesota
Technical Report No. 611

January, 1996

¹ Konrad Zuse Center for Information Technology, D-10711 Berlin, Germany.

Bias corrected bootstrap bandwidth selection

Birgit Grund* and Jörg Polzehl†

January 18, 1996

Abstract

Current bandwidth selectors for kernel density estimation that are asymptotically optimal often prove less promising under more moderate sample sizes. The point of this paper is to derive a class of bandwidth selectors that attain optimal root- n convergence while still showing good results under small and moderate sample sizes. This is achieved by minimizing bias corrected smoothed bootstrap estimates of the mean integrated squared error. The degree of bias correction determines the rate of relative convergence of the resulting bandwidth. The bias correction targets finite sample bias rather than asymptotically leading terms, resulting in substantial improvements under smaller sample sizes. In a simulation study, the new methods are compared to several of the currently most popular bandwidth selectors, including plug-in, cross-validation and other bootstrap rules. Practical implementation, and the choice of the oversmoothing bandwidth are discussed.

Keywords: Smoothed bootstrap, bias correction, bandwidth selection,
kernel density estimation.

AMS-Classification: Primary 62G05, Secondary 62G25

*School of Statistics, University of Minnesota, St. Paul, MN 55108, U.S.A.

†Konrad Zuse Center for Information Technology, D-10711 Berlin, Germany.

1 Introduction

A central problem in kernel density estimation is the data-driven selection of smoothing parameters. During the recent years, many different methods have been proposed. For an overview see Jones, Marron and Sheather (1992); the need for data-driven bandwidth selection is motivated in Silverman (1986), among others.

There are two major approaches to deriving bandwidth selectors; (i) to minimize some empirical criterion that measures goodness-of-fit, and (ii) bandwidth selectors that aim at optimal convergence rates, and minimize asymptotically leading terms of the risk. The first type includes cross-validation (CV) methods and bootstrap bandwidth selection, as proposed in Taylor (1989) and Faraway and Jhun (1990). The second type comprises a large variety of bandwidth selectors, including the popular plug-in bandwidth selectors proposed in Park and Marron (1990), Hall, Sheather, Jones and Marron (1991), Jones, Marron and Park (1991), etc. A good overview of these methods is given in Jones, Marron and Sheather (1992).

Unfortunately, bandwidth selectors that are designed to achieve optimal convergence rates tend to be less favorable for moderate sample sizes, or, citing Jones, Marron and Sheather (1992), "...achieving optimal theoretical performance (up to the bounds derived by Hall and Marron, 1991) and acceptable practical performance is not accomplished by the same techniques." The purpose of this paper is to derive a data-driven bandwidth selector that achieves both, the fastest possible theoretical rate of relative convergence, $n^{-1/2}$, and small sample behavior that competes with the best of the global bandwidth selectors on the market.

Let X_1, \dots, X_n denote a random sample from a density f . The kernel density estimate \hat{f}_h estimates the curve f pointwise by

$$\hat{f}_h(x) = n^{-1} \sum_{i=1}^n K_h(x - X_i),$$

where $K_h(x) = h^{-1}K(x/h)$, with some kernel function K and some bandwidth h . In order to develop our bootstrap bandwidth selector and to derive theoretical properties, we consider as "optimal" bandwidth the minimizer h_0 of the mean integrated squared error (*MISE*),

$$MISE_f(h) = \mathbf{E} \int (\hat{f}_h(x) - f(x))^2 dx.$$

However, in simulation studies we use the the integrated squared error (*ISE*) to compare bandwidth selectors, in order to properly reflect the behavior of the bandwidth procedures for the *sample at hand*. A justification for this approach is given in Grund, Hall and Marron (1994) and Jones (1991).

We introduce a class of bandwidth selectors that is based on smoothed bootstrap estimates for the mean integrated squared error with additive bias correction. The repeated

application of our bias correction principle leads to bandwidth selectors with increasingly fast rates of convergence. Simple bias correction allows for a convergence rate of $n^{-4/13}$, the same rate as obtained by Park and Marron (1990). Additional bias correction steps lead to convergence rates of $n^{-8/17}$ and $n^{-1/2}$, respectively. These rates have been obtained by Jones, Marron and Park (1991) by minimizing variations of the smoothed bootstrap criterion, and choosing g optimally in the form of $g \sim n^p h^m$ ($m = -2$ for non-negative kernel). Note that $n^{-1/2}$ is the best possible rate, in the sense of Hall and Marron (1991). In defining the bootstrap, we use the theoretical expectation over the re-sampling space. Therefore, we provide explicit formulae for computing the bootstrap bandwidth selector and avoid actual re-sampling in computing the bandwidth for a sample at hand.

The bias correction procedure is described in detail in Section 2; this section as well contains a discussion of the smoothed bootstrap. Theoretical results are given in Section 3. Proofs are deferred to Section 5. Section 4 is dedicated to a simulation study. Here, the new bandwidth selectors are compared to several of the currently most popular bandwidth procedures, including several plug-in methods, least squares cross-validation, and Faraway and Jhun's empirical bootstrap. The simulations are based on the 15 normal mixture densities introduced in Marron and Wand (1992).

In this paper, we restrict our considerations to density estimation, and to global bandwidth choice. However, it can be expected that the here introduced techniques are useful for developing bandwidth selectors in a variety of settings, including regression, hazard and survival curve estimation, as well as for local bandwidth procedures.

2 The bandwidth selector

2.1 Smoothed bootstrap

Bootstrap procedures for selecting the bandwidth in kernel density estimation have been first proposed by Taylor (1989) and Faraway and Jhun (1990). In the latter paper, the bootstrap bandwidth \hat{h}_{FJ} is defined as the minimizer of a smoothed bootstrap estimate of MISE,

$$\widehat{MISE}_f(h) = MISE_{\hat{f}_g}(h) = \mathbf{E}^* \int (\hat{f}_h^* - \hat{f}_g)^2. \quad (2.1)$$

Here, \hat{f}_g is a pilot kernel estimate for the unknown target density f ; \hat{f}_h^* denotes the kernel estimate based on a bootstrap sample (of size n) drawn from the pilot density \hat{f}_g , and \mathbf{E}^* denotes the expectation over the bootstrap samples.

Note that the theoretical expectation \mathbf{E}^* over the bootstrap samples can be expressed as a random variable in X_1, \dots, X_n ,

$$MISE_{\hat{f}_g}(h) = \int (K_h * \hat{f}_g - \hat{f}_g)^2 + n^{-1} \int K_h^2 - n^{-1} \int (K_h * \hat{f}_g)^2, \quad (2.2)$$

due to the simple form of \hat{f}_g . Here, $*$ denotes the convolution. The first term in (2.2) stands for the integrated squared bias, the other two terms are the integrated variance. Throughout this paper we use the *theoretical* expectation over the bootstrap samples (in this case, the expectation with respect to density \hat{f}_g , conditional on X_1, \dots, X_n) when we refer to the bootstrap criteria. Originally, Faraway and Jhun (1990) defined their bootstrap criterion through the empirical mean, i.e., \mathbf{E}^* would denote the average over a certain number of bootstrap resamples.

Bandwidth selection criteria of the form (2.1) have been investigated in Hall, Marron and Park (1989) and Jones, Marron and Park (1991) under the name of “smoothed cross-validation”. The first paper suggests to employ higher order kernels L for estimating the pilot density \hat{f}_g , and discusses optimal choice of L and g . In contrast, Jones, Marron and Park (1991) achieve the optimal convergence rate of $(\hat{h} - h_0)/h_0 \sim n^{-1/2}$ already with second order kernels, by tying the pilot bandwidth g to h in the form of

$$g = Cn^p h^m,$$

for some constants C, p and $m = -2$.

It is well known that the bandwidth g of the pilot estimate strongly influences the behavior of the bootstrap; see Hall, DiCiccio and Romano (1989) and Silverman and Young (1987). There is a general understanding that, asymptotically, g has to be chosen an order of magnitude larger than h in order to improve on the convergence rate of $(\hat{h} - h_0)/h_0 \sim n^{-1/10}$. (Therefore, we refer to g as “oversmoothing bandwidth.” For justification and proofs in the context of smoothed bootstrap, see Hall, Marron and Park, 1992, and Jones, Marron and Park, 1991). However, appropriate data-dependent choice of g beyond adjusting convergence rates remains a difficult problem.

Faraway and Jhun (1990) suggest selecting an initial value of g through least-squares cross-validation, and possibly iterating with $g_{i+1} = h_i$. Simulation studies show that the resulting density estimate computed with \hat{h}_{FJ} is usually satisfactory, with risks comparable to other, well-accepted bandwidth rules. However, the relative convergence rate of \hat{h}_{FJ} is only $n^{-1/10}$, due to $g \sim n^{-1/5}$.

In the current paper, we provide the asymptotically optimal g for our estimates in Section 3. The data-dependent choice of g is discussed in Section 4.

2.2 Motivating bias corrected bootstrap

The bootstrap criterion (2.1) as an estimate of $MISE_f(h)$ is biased. Moreover, the mean squared error of $MISE_{\hat{f}_g}(h)$ is asymptotically dominated by the bias term,

$$\mathbf{E}MISE_{\hat{f}_g}(h) - MISE_f(h) = \mathbf{E}R(D_h * \hat{f}_g) - R(D_h * f) - \frac{1}{n}\mathbf{E}\{R(K_h * \hat{f}_g) - R(K_h * f)\}, \quad (2.3)$$

where D_h and R are defined by $D_h * \psi = K_h * \psi - \psi$ and $R(\psi) = \int (\psi)^2$, respectively, for any function ψ . Therefore, asymptotically optimal bandwidth choice has to target primarily the bias. The root- n bandwidth procedures of Hall, Marron and Park (1989) and Jones, Marron and Park (1991) reduce the bias by eliminating asymptotically leading terms in a Taylor expansion of $\mathbf{E}(\hat{h} - h_0)^2$, either by employing higher order kernels, or by carefully balancing h and g . In practical implementation, some constants that depend on the unknown density f have to be estimated. Conceptually, the final bandwidth h is still chosen by minimizing the biased bootstrap criterion (2.2). The fine-tuning through the pilot bandwidth g targets only *asymptotically* dominant bias terms. For any densities where trailing terms in $\mathbf{E}(\hat{h} - h_0)^2$ are negligible only for rather large n , these methods can be expected to fail under moderate sample sizes, even given the most sophisticated estimation of g . (This problem is discussed in more detail in Marron and Wand (1992). Among others, the authors provide striking figures that illustrate differences between the true MISE and the asymptotic part of MISE that is used for root- n bandwidth selection for various densities.)

In contrast, we suggest to implement a bias correction for the bootstrap criterion $MISE_{\hat{f}_g}(h)$ itself. More important, we estimate the complete finite sample bias (2.3) by bootstrap methods, rather than focusing on asymptotically leading terms. Repeated application of the bias correction procedure results in classes of bandwidth selectors with increasing rates of relative convergence, including the fastest possible speed of root- n . Our simulation study in Section 4 shows that these “bias corrected” root- n bandwidth selectors are comparable to the root- n estimate of Park, Marron and Jones (1991) for large sample sizes, while behaving favorably for smaller and moderate samples as well.

The bias correction procedure is described in detail in the next section.

2.3 Bias corrected bootstrap

In order to obtain root- n convergence, several steps of bias correction are necessary. Note that the bootstrap criterion $MISE_{\hat{f}_g}(h)$ can be written as

$$MISE_{\hat{f}_g}(h) = R(D_h * \hat{f}_g) - n^{-1}R(K_h * \hat{f}_g) + n^{-1}R(K_h).$$

Since K is symmetric, the mean of the first two terms can be represented through

$$\mathbf{E}R(D_h * \hat{f}_g) = \frac{n-1}{n}R(D_h * K_g * f) + \frac{1}{n}R(D_h * K_g),$$

and

$$\mathbf{E}R(K_h * \hat{f}_g) = \frac{n-1}{n}R(K_h * K_g * f) + \frac{1}{n}R(K_h * K_g),$$

respectively. Consequently, the criterion

$$\begin{aligned}\widetilde{MISE}_{\hat{f}_g}(h) = & \frac{n}{n-1}R(D_h * \hat{f}_g) - \frac{1}{n-1}R(D_h * K_g) + \frac{1}{n}R(K_h) \\ & - \frac{1}{n-1}R(K_h * \hat{f}_g) + \frac{1}{n(n-1)}R(K_h * K_g)\end{aligned}\quad (2.4)$$

is an unbiased estimate for $MISE_{K_g * f}$. For the sake of brevity we omit regularity conditions here; details are given in Section 3.

The bandwidth selectors investigated in the current paper are all based on the initial criterion $\widetilde{MISE}_{\hat{f}_g}(h)$. Note that $\widetilde{MISE}_{\hat{f}_g}$ is unbiased only for estimating $MISE_{K_g * f}$ rather than for the target criterion $MISE_f$. Therefore, we suggest to estimate the remaining bias, $MISE_{K_g * f} - MISE_f$, by $\widetilde{MISE}_{K_g * \hat{f}_g} - \widetilde{MISE}_{\hat{f}_g}$. Repeating this process leads to the following *extrapolation criteria*:

$$\hat{M}_0(h) = \hat{M}_0(h; \hat{f}_g) = \widetilde{MISE}_{\hat{f}_g}, \quad (2.5)$$

$$\begin{aligned}\hat{M}_1(h) &= \hat{M}_1(h; \hat{f}_g) = \hat{M}_0(h; \hat{f}_g) - (\hat{M}_0(h; K_g * \hat{f}_g) - \hat{M}_0(h; \hat{f}_g)) \\ &= 2\widetilde{MISE}_{\hat{f}_g} - \widetilde{MISE}_{K_g * \hat{f}_g},\end{aligned}\quad (2.6)$$

$$\begin{aligned}\hat{M}_2(h) &= \hat{M}_2(h; \hat{f}_g) = \hat{M}_1(h; \hat{f}_g) - (\hat{M}_1(h; K_g * \hat{f}_g) - \hat{M}_1(h; \hat{f}_g)) \\ &= 3\widetilde{MISE}_{\hat{f}_g} - 3\widetilde{MISE}_{K_g * \hat{f}_g} + \widetilde{MISE}_{K_g * K_g * \hat{f}_g}.\end{aligned}\quad (2.7)$$

The corresponding bandwidth selectors \hat{h}_κ , $\kappa = 0, 1, 2$, are defined as minimizers of the criteria $\hat{M}_\kappa(h)$.

Note that $K_g * \hat{f}_g$ is simply the kernel estimate based on $K_g * K_g = (K * K)_g$. Therefore, $\widetilde{MISE}_{K_g * \hat{f}_g}$ is unbiased for $MISE_{K_g * K_g * f}$. Consequently, formulae (2.6) and (2.7) allow one to interpret \hat{M}_1 and \hat{M}_2 as linear and quadratic extrapolation, respectively, to a criterion that is unbiased for estimating $MISE_f$.

In a second step, we suggest estimating the remaining bias through a reference density η . Here, mixtures of normal distributions are a convenient choice, since they are flexible enough to track a wide variety of density patterns and also have useful theoretical properties. We define the following *normal reference criteria*:

$$\hat{N}_0(h) = \hat{N}_0(h; \hat{f}_g) = \hat{M}_0(h; \hat{f}_g) - (\mathbf{E}\hat{M}_0(h; \hat{\eta}_g) - MISE_\eta(h)),$$

$$\hat{N}_1(h) = \hat{N}_1(h; \hat{f}_g) = \hat{M}_1(h; \hat{f}_g) - (\mathbf{E}\hat{M}_1(h; \hat{\eta}_g) - MISE_\eta(h)),$$

$$\hat{N}_2(h) = \hat{N}_2(h; \hat{f}_g) = \hat{M}_2(h; \hat{f}_g) - (\mathbf{E}\hat{M}_2(h; \hat{\eta}_g) - MISE_\eta(h)),$$

where η is a mixture of normal densities, and $\hat{\eta}_g$ denotes the kernel estimate with bandwidth g applied to a sample from the reference distribution η . In practical application, the parameters of η will be estimated. For example, in case of $\eta = \phi_{\hat{\sigma}}$, the Gaussian density, we suggest using $\hat{\sigma} = \min(IQR/1.34, s)$, where IQR denotes the interquartile range, and s the sample standard deviation.

Obviously, the concept of bias correction via reference densities is not limited to the normal distribution. However, in the special case when K is the Gaussian kernel and η is a normal density or a mixture of normal densities, closed form formulae for the exact values of $\mathbf{E}\hat{M}_\kappa(h; \hat{\eta}_g)$, $\kappa = 0, 1, 2$, and $MISE_{\hat{\eta}_g}$ are available (Marron and Wand, 1992). For practical implementation, we therefore favor a Gaussian kernel and normal mixture reference distributions.

The linear and quadratic reference criteria \hat{N}_1 and \hat{N}_2 combine the extrapolation and the normal reference approach to bias correction. The influence of the normal reference distribution is least for the quadratic criterion.

Similar to ordinary smoothed bootstrap, the “bias corrected bootstrap bandwidth selectors” defined here strongly depend on the bandwidth g . Theoretically optimal choice of g is discussed in Section 3, while computational implementation is described in Section 4.

3 Asymptotic theory

Let $\mu_i(K) = \int x^i K(x) dx$ denote the i -th moment of K . The following assumptions are used:

(A1) K is a symmetric probability density, yielding $\mu_{10}(K) < \infty$.

(A2) K has ten bounded derivatives, $K^{(10)}$ is Hölder continuous, and

$$\sup_{x \in \mathbb{R}} |x^2 K''(x)| < \infty.$$

(A3) f has twelve bounded derivatives, $f^{(12)}$ is Hölder continuous, and

$$\sup_{x \in \mathbb{R}} |x^2 f''(x)| < \infty.$$

Let \hat{h}_κ , $\kappa = 0, 1, 2$, denote the minimizer of \hat{M}_κ or \hat{N}_κ (as needed), minimizing over some interval (ϵ_n, ∞) , with $\epsilon_n \rightarrow 0$, $n \epsilon_n \rightarrow \infty$. Note that \hat{N}_κ contains additional bias correction with a reference density η .

Theorem 3.1 Assume that $h_0/g \rightarrow 0$, $g \rightarrow 0$, $ng^2 \rightarrow \infty$, and that (A1) – (A3) hold. Then,

$$(i) \quad (\hat{h}_\kappa - h_0)/h_0 = n^{2/5} h_0^2 \left\{ \left[n^{-2} g^{-9} C_{\sigma,1} + n^{-1} C_{\sigma,2} \right]^{1/2} Z_n + g^{2+2\kappa} C_{\mu,\kappa,\eta} \right\}$$

for $\kappa = 0, 1, 2$, where Z_n is asymptotically normal $N(0,1)$, and

$$C_{\mu,\kappa,\eta} = -\mu_2(K)^{3+\kappa} C_{f,\eta} C_D^{-1},$$

$$C_{f,\eta} = \begin{cases} R(f^{(3+\kappa)}), & \text{if } \hat{h}_\kappa \text{ minimizes } \hat{M}_\kappa, \\ R(f^{(3+\kappa)}) - R(\eta^{(3+\kappa)}), & \text{if } \hat{h}_\kappa \text{ minimizes } \hat{N}_\kappa; \end{cases}$$

$$C_D = 5R(f'')^{3/5} R(K)^{2/5} \mu_2(K)^{6/5},$$

$$C_{1,\sigma} = 2\mu_2(K)^4 R(f) R(K * K^{(4)}) C_D^{-2},$$

$$C_{2,\sigma} = 4\mu_2(K)^4 \left[\int (f^{(4)})^2 f - R(f'')^2 \right] C_D^{-2}.$$

(ii) $E[\hat{h}_\kappa - h_0]^2$ is minimized when

$$g_\kappa = n^{-2/(13+4\kappa)} G_\kappa \{1 + o(1)\} \quad \text{for } \kappa = 0, 1, 2,$$

where

$$G_\kappa = \left[\frac{9}{4 + 4\kappa} C_{1,\sigma} C_{\mu,\kappa,\eta}^{-2} \right]^{1/(13+4\kappa)}.$$

Corollary 3.2 When $g \sim n^{-2/(13+4\kappa)}$, then the rate of relative convergence of \hat{h}_κ to h_0 yields:

$$(\hat{h}_\kappa - h_0)/h_0 \sim \begin{cases} n^{-(4+4\kappa)/(13+4\kappa)} & \text{for } \kappa = 0, 1, \\ n^{-1/2} & \text{for } \kappa = 2. \end{cases}$$

Remark 3.3 All proofs are deferred to Section 5. Corollary 3.2 follows immediately from Theorem 3.1.

Remark 3.4 The assumptions of Theorem 3.1 could be weakened at the expense of slightly longer proofs. In particular, we need only ten derivatives of f . Additionally, Theorem 3.1 (i) requires only $ng^{3/2} \rightarrow \infty$ instead of $ng^2 \rightarrow \infty$.

Remark 3.5 Corollary 3.2 implies that root- n convergence is obtained only for $\kappa = 2$. This means that at least quadratic extrapolation is required to sufficiently reduce the bias of the bootstrap criterion in order to achieve optimal relative convergence. For non-negative

kernels, this rate was first achieved by Jones, Marron and Park (1991). Here, bandwidth choice was based on the uncorrected $MISE_{\hat{f}_g}$, with the oversmoothing parameter $g = Cn^p\hat{h}^m$ chosen to minimize asymptotically leading terms of $\mathbf{E}[\hat{h} - h_0]^2$. In contrast, we propose estimating first the finite sample bias of $MISE_{\hat{f}_g}$ (not just asymptotically leading terms), and only then to fine-tune g .

Our simulation study shows similar behavior of both bandwidth selectors for large sample size ($n = 1600$), in tendency favoring our estimate. For small and medium sample sizes, the relative performance depends on the structure of the underlying density. For unimodal densities, our estimate is often substantially better, whereas for bimodal densities the root- n estimate of Jones, Marron and Park (1991) is superior. This effect cannot be simply attributed to the use of the normal density as reference for the additional bias correction in \hat{N}_κ , since it occurs for both the \hat{N}_κ and \hat{M}_κ rules. (Simulations show that choosing the reference density η to roughly follow the structure of the data results in substantial improvements of \hat{N}_κ for those densities.) The effect wears off with increasing sample size. For details see Section 4.

Remark 3.6 In the case of simple bias correction ($\kappa = 0$), the rate of relative convergence is $n^{-4/13}$. This is the rate obtained by the plug-in bandwidth selector of Marron and Park (1990).

Linear extrapolation ($\kappa = 1$) leads to a bandwidth selector with a relative convergence rate of $n^{-8/17}$. Jones, Marron and Park (1991) obtained the same rate by minimizing $MISE_{\hat{f}_g}$, again using a factorization $g = Cn^{-23/45}h^{-2}$, and leaving out diagonal elements in optimizing the integrated squared bias part.

Remark 3.7 Our additional bias correction through the normal reference procedure in \hat{N}_κ does not change the theoretical rate of convergence. However, it does change the constant $C_{\mu,\kappa,\eta}$. This accounts for the (sometimes substantial) improvements in finite sample behavior over a wide variety of densities, as they show in the simulation study in Section 4. While improvements are most marked when f is close to normal (as can be expected), the additional normal reference bias correction seems not to have adverse effects even when f is strongly skewed or markedly multimodal.

4 Simulation

In this section we report details of a simulation study carried out to evaluate the performance of our bandwidth selection procedures. The study is based on the 15 normal mixture densities given in Marron and Wand (1992). One advantage of these densities is that we can calculate $MISE_f(h)$ explicitly.

For each density, we generated 500 samples of size $n = 100, 400$ and 1600 (which may be considered as small, medium and large sample sizes). All density estimates are computed using a Gaussian kernel, K , and Fast Fourier Transformation based on 2^9 points. Whenever necessary, we use $\hat{\sigma} = \min(IQR/1.34, s)$ to estimate the scale.

We compare the bias corrected bootstrap criteria with several well known bandwidth selectors:

SJ: The method proposed by Sheater and Jones (1991)¹. The selected bandwidth \hat{h}_{SJ} is computed as solution of

$$h = \left(\frac{R(K)}{\sigma_K^4 R(\hat{f}_{g(h)}'')} \right)^{1/5} n^{-1/5}$$

with

$$g(h) = 1.357 \left(\frac{R(\hat{f}_a'')}{R(\hat{f}_b''')} \right)^{1/7} h^{5/7}$$

$$a = 0.92 IQR n^{-1/7} \quad \text{and} \quad b = 0.912 IQR n^{-1/9}.$$

RN: The root- n bandwidth selector \hat{h}_{RN} with second order normal reference according to Jones, Marron and Park (1991). The selected bandwidth is the minimizer of

$$\int (K_h * \hat{f}_g - \hat{f}_g)^2 + (nh)^{-1} R(K)$$

where g depends on h in the form of

$$g = 0.6297 R(\hat{f}_a'')^{-2/5} R(\hat{f}_b^{IV})^{-1/9} n^{-23/45} h^{-2}. \quad (4.1)$$

The bandwidths a and b are based on a normal reference and an initial bandwidth selected by the method of Sheather and Jones (1991)

$$a = .91 \hat{h}_{SJ} n^{3/65} \quad \text{and} \quad b = .86 \hat{h}_{SJ} n^{7/85}.$$

The constants in (4.1) are chosen to give best results in estimating the roughness terms for $n = 1000$ and normal density.

PM: The bandwidth selector proposed in Park and Marron (1990). The selected bandwidth \hat{h}_{PM} is the solution of the equation

$$h = \left(\frac{R(K)}{\sigma_K^4 [R(\hat{f}_{g(h)}'') - n^{-1} g(h)^{-5} R(K'')]} \right)^{1/5} n^{-1/5},$$

¹the routines used are courtesy of S.J. Sheather

with $g(h)$ obtained as

$$g(h) = 0.8865 \hat{\sigma}^{3/13} h^{10/13}$$

using a normal reference for the constant and $\hat{\sigma} = \min(IQR/1.34, s)$.

LSCV: The bandwidth selector minimizing the Least Squares Cross-Validation criterion

$$LSCV(h) = R(\hat{f}_h) - 2n^{-1} \sum_{i=1}^n \hat{f}_{h;-i}(x_i)$$

introduced by Rudemo (1982) and Bowman (1984). Here, $\hat{f}_{h;-i}$ is the density estimate computed without use of the i -th observation. The bandwidth \hat{h}_{LSCV} is computed minimizing the closely related criterion $M_1(h)$ given in Silverman (1986). The minimum is searched in two steps, over logarithmically spaced grids. The coarser grid consists of 76 points ranging from $\frac{1}{15}\hat{h}_{RT}$ to $2\hat{h}_{RT}$, with $\hat{h}_{RT} = 1.06\hat{\sigma}n^{-1/5}$ being Silverman's "Rule of Thumb" bandwidth selector. The finer grid includes 101 points, centered on the minimum found in the first step.

FJ: The smoothed bootstrap bandwidth selector proposed in Faraway and Jhun (1990). Here we use the theoretical mean. The oversmoothing bandwidth g is chosen iteratively, $g_{i+1} = h_i$, with pilot value $g = \hat{h}_{LSCV}$. This iteration dampens the influence of the initial choice of g , and improves the stability of the final estimate.

Additionally, we include the bandwidth \hat{h}_{ISE} minimizing

$$ISE(h) = \int (\hat{f}_h - f)^2$$

in our comparison. Note that \hat{h}_{ISE} is the theoretically optimal choice for the sample at hand.

The bias corrected bootstrap procedures are described in Section 2.3. Similar to the first two plug-in estimates, the choice of the oversmoothing bandwidth g requires some care. We target the optimal bandwidth g_κ , as given in Theorem 3.1 (ii) for \hat{M}_κ . Since g_κ depends on the unknown f , we use

$$\hat{g}_\kappa = n^{-2/(13+4\kappa)} \left(\frac{9R(K * K^{(4)})R(\hat{f}_a)}{(2 + 2\kappa)R(\hat{f}_{b_\kappa}^{(3+\kappa)})^2} \right)^{1/(13+4\kappa)}, \quad (4.2)$$

replacing f by suitable estimates \hat{f}_a and \hat{f}_b . The bandwidths a and b are aimed to approximate optimal oversmoothing bandwidths for estimating $R(f)$ and $R(f^{(3+\kappa)})$, respectively.

They are defined by

$$\begin{aligned}
a &= .9224n^{-1/5}h, \\
b_0 &= \left(1.165R(\hat{f}_a)/R(\hat{f}_{d_0}^{(4)})^2\right)^{1/17} R(\hat{f}_c''/.282)^{1/5}n^{7/85}h, \\
b_1 &= \left(241.9R(\hat{f}_a)/R(\hat{f}_{d_1}^{(5)})^2\right)^{1/21} R(\hat{f}_c''/.282)^{1/5}n^{11/105}h, \\
b_2 &= \left(2425R(\hat{f}_a)/R(\hat{f}_{d_2}^{(6)})^2\right)^{1/25} R(\hat{f}_c''/.282)^{1/5}n^{3/25}h,
\end{aligned} \tag{4.3}$$

where

$$\begin{aligned}
c &= .7486n^{3/65}h, & d_0 &= .9433n^{11/105}h, \\
d_1 &= .9027n^{3/25}h, & \text{and } d_2 &= .9083n^{19/145}h.
\end{aligned} \tag{4.4}$$

The bandwidths a , b_κ , c and d_κ are plug-in estimates of the form $Cn^{-p}h$. Here, c and d_κ are chosen to “optimally” estimate $R(f'')$ and $R(f^{(4+\kappa)})$, respectively, optimal for a normal reference density $f = \eta_\sigma$. The bandwidth selectors \hat{h}_κ are computed iteratively, as follows:

Start: Pilot value $\hat{h}_{\kappa,0} = \hat{h}_{LSCV}$.

- Step i :**
- Compute bandwidth a , b_κ , c and d_κ as given in (4.3) and (4.4) with $h = \hat{h}_{\kappa,i}$, and $\hat{g}_{\kappa,i}$ as in (4.2).
 - Choose $\hat{h}_{\kappa,(i+1)}$ to minimize the criterion \hat{M}_κ (or \hat{N}_κ) with $g = \hat{g}_{\kappa,i}$; set $i := i + 1$.

Stop: When $\hat{h}_{\kappa,i}$ stabilizes.

Tables 1–3 display averages of $\left((\hat{h} - h_0)/h_0\right)^2$, $\left(ISE(\hat{h}) - MISE(h_0)\right)/MISE(h_0)$ and $\left(MISE(\hat{h}) - MISE(h_0)\right)/MISE(h_0)$, respectively, for each of the 12 bandwidth selection criteria, ISE , RN , SJ , PM , $LSCV$, FJ , \hat{M}_κ and \hat{N}_κ ($\kappa = 0, 1, 2$), and the 15 normal mixture densities in Marron and Wand (1992). In order to assess the variability, we additionally provide the minimal and maximal values of the coefficients of variation over the 12 bandwidth selectors, for each of the densities. The values in Table 1 measure the behavior of the bandwidths as estimates of h_0 . Table 2 describes the relative performance of the \hat{h} in estimating the density *for the sample at hand* (using ISE), while the results in Table 3 assess the relative performance with respect to $MISE$.

All three tables show that our bias corrected bootstrap procedures behave quite well for the unimodal densities (# 1-5), in particular the \hat{N}_2 rules. Here, the additional bias correction via normal reference leads to often substantial improvements of the \hat{N}_κ rules over the \hat{M}_κ rules for small and medium sample sizes. These improvements are strongest for $\kappa = 0$ (simple bias correction), and decrease with increasing sample size.

Density # 1 (normal density) is of particular interest, since the bias correction in the \hat{N}_κ rules is performed using the true underlying density. Note that the risk here is increasing from \hat{h}_{N_0} through \hat{h}_{N_2} . This phenomenon clearly reflects the additional variability introduced by subsequent bias correction steps. For all sample sizes, the \hat{h}_{M_1} and \hat{h}_{M_2} values are better than \hat{h}_{RN} , \hat{h}_{SJ} , \hat{h}_{PM} and \hat{h}_{LSCV} , competing only with \hat{h}_{FJ} . As opposed to the \hat{N}_κ rules, the \hat{M}_κ rules do not take undue advantage of the true density for bias correction; obviously the \hat{N}_κ rules are favored by density # 1.

For the multimodal densities # 6-9 the bandwidth selectors \hat{h}_{SJ} and \hat{h}_{RN} are superior (for small and large sample sizes, respectively). Our procedure \hat{N}_2 follows on place 2 or 3. Additional simulation studies have shown strong improvements of the \hat{N}_κ using a bivariate reference density η . Here, \hat{N}_0 and \hat{N}_1 were superior for low and moderate sample size, and \hat{N}_2 for $n = 1600$.

For density # 3 (with small sample size) and densities #10-15, *LSCV* performs best. This is due to the bias introduced in all other rules by the selection of reference bandwidths at different stages and the only *asymptotic* optimality of the plug-in rules. Again, the \hat{N}_2 rules range on places 2 and 3. The results obtained for these densities indicate that our procedures are rather adaptive in complicated situations.

Note that \hat{h}_{FJ} performs surprisingly well across most of the densities, although this bandwidth selector does not contain explicit bias correction.

We selected four of the densities for a more detailed analysis. We decided on the standard normal density (which is used as a reference density for most of the bandwidth selectors in one or the other way), the strongly skewed density # 3, the claw density # 10 and the asymmetric bimodal density #8. The last density, #8, was chosen because it gave the least favorable results for \hat{h}_{M_2} and \hat{h}_{N_2} . Figures 1, 4, 6 and 8 show boxplots of values $ISE(h)$ obtained from 500 simulated samples for the three sample sizes. The vertical scale has been adjusted (truncating a small percentage of the values larger than $C \cdot n^{-4/5}$) in order to make the results for different sample sizes more comparable. Boxplots of the corresponding $(\ln \hat{h} - \ln \hat{h}_{ISE})$ values are displayed in Figures 2, 5, 7 and 9.

Figures 1 and 2 demonstrate superior performance of \hat{N}_2 , although all other selectors except *LSCV* also perform reasonably well in this situation. We included Figure 3 to illustrate the effect of the normal reference bias correction involved in \hat{h}_{N_2} . Figures 4 and 5 indicate that, compared to the plug-in rules, \hat{h}_{M_2} and \hat{h}_{N_2} are better handling the bias introduced by the skewness of the density.

The Claw density # 10 was used in Marron and Wand (1992) to illustrate the existence of multiple minima of $MISE(h)$ (for sample size $n = 53$). Figures 6 and 7 reflect that \hat{h}_{M_2} and \hat{h}_{N_2} find the correct location of the global minimum for at least part of the samples for $n = 100$, and almost always for $n = 400$, while the plug-in criteria fail in most of

the cases.

Figures 8 and 9 display the corresponding plots for density # 8 which was the least favorable for our proposals. For $n = 100$ and $n = 400$ the plug-in selectors RN and SJ are superior, but even here \hat{h}_{M_2} and \hat{h}_{N_2} perform better than \hat{h}_{PM} , \hat{h}_{LSCV} and \hat{h}_{FJ} . For $n = 1600$ the root- n convergence takes effect, and \hat{h}_{M_2} and \hat{h}_{N_2} are close to \hat{h}_{RN} and better than \hat{h}_{SJ} . It is to be noted that throughout all 15 densities the \hat{N}_κ rules have smaller or equal risk than the corresponding \hat{M}_κ rules. In tendency, \hat{h}_{M_2} and \hat{h}_{N_2} perform better than \hat{h}_{M_1} , \hat{h}_{N_1} , \hat{h}_{M_0} and \hat{h}_{N_0} , with the effect being stronger for the \hat{M}_κ rules.

Choice of g : We decided to use the oversmoothing bandwidth g as defined by (4.2) - (4.4). This choice provides the rates of convergence stated in Corollary 3.2.

Obviously, there is some arbitrariness in taking a, b_κ, c and d_κ proportional to h . We also considered choosing a, b_κ, c and d_κ independently of h . In this case, the final estimates can be computed straightforwardly, without iterating h . However, this leads mostly to larger values of \hat{g}_κ , and to higher risks. Even so, the choice of g seems to be far from optimal for smaller sample sizes .

Using $g := h$ instead of (4.2) in the iteration, we often obtained better results for the \hat{M}_2 and \hat{N}_2 -rules; under $n = 100$ for densities # 3, 4, 6, 8, 9, 11 and 13-15; under $n = 400$ for densities # 3, 4 and 8-10, 12-15 and under $n = 1600$ for densities # 3, 9 and 12-15. On the other hand side, using \hat{g}_κ leads to significantly better results in case of the densities # 1, 2, 5, 7 and 11, especially for larger sample sizes. Here, the relative convergence rate of \hat{h}_κ clearly takes effect. In most cases, the iteration $g := h$ stabilized at smaller values of \hat{g} . This suggests that our \hat{g}_κ tends to overestimate the optimal oversmoothing bandwidth g for more structured densities. In case of densities # 3, 4, 6, 8, 9 and 11 the effect was stronger for $n = 100$. For the “difficult to estimate” densities # 12-15 even a sample size of $n = 1600$ seems to be far from the point where asymptotics start to work.

Note that the choice of $g := h$ results in the rather slow convergence $(\hat{h} - h_0)/h_0 \sim n^{-1/10}$, while \hat{g}_κ for $\kappa = 2$ gives the best possible root- n rate. Again, this clearly shows that an oversmoothing bandwidth that allows a higher rate of relative convergence for \hat{h} does not necessarily result in better performance in practical applications, in particular when working with moderate sample sizes.

Choice of η : The use of an appropriate normal mixture η as reference density in \hat{N}_κ often considerably improves results. The density η might be chosen by prescreening and selecting a normal mixture that roughly follows the structure of the data.

Additionally to $\eta = \phi$, we implemented the mixture of two normal densities, $\eta = \omega\phi_{\mu_1, \sigma_1} + (1 - \omega)\phi_{\mu_2, \sigma_2}$. The parameters ω, μ_i and $\sigma_i, i = 1, 2$, were estimated by maxi-

mum likelihood. Initial values were obtained clustering the observations and taking $\omega = .5$ and μ_i, σ_i as empirical moments of the clusters. We denote the resulting rules by $\hat{N}_{\kappa,B}$.

The results obtained are presented in Tables 4 - 6. The $\hat{N}_{\kappa,B}$ rules are less efficient than the \hat{M}_κ and \hat{N}_κ rules only in case of the normal density (# 1), and density # 2 with $n = 100$, while significant gains were observed for most other situations.

Summary: Overall, our preferred bandwidth choice is minimizing the highest order normal reference criterion, \hat{N}_2 . The resulting estimate combines quite favorable small sample size behavior with optimal root- n convergence. In some cases, \hat{h}_{M_2} and \hat{h}_{N_2} turn out to be considerably better than most of the other investigated methods, then competing only with \hat{h}_{RN} for large sample sizes. While in certain settings $\hat{h}_{RN}, \hat{h}_{SJ}, \hat{h}_{FJ}$ or \hat{h}_{LSCV} have smaller risk, the bias corrected estimates never break down completely for any of the 15 densities. Their risks are either substantially smaller, or close to the median risk of the other estimates. In contrast, $\hat{h}_{RN}, \hat{h}_{SJ}, \hat{h}_{FJ}, \hat{h}_{PM}$ and \hat{h}_{LSCV} each fail for at least two of the 15 densities. Additionally, results for the \hat{N}_κ rules can even be considerably improved for small and medium sample sizes using an appropriately selected reference density η .

5 Proofs

For the sake of brevity, we give detailed proofs only for the extrapolation rules \hat{M}_κ , with short remarks on the normal reference rules \hat{N}_κ . Here and below, we write $\hat{M}_\kappa, \hat{N}_\kappa$ and \hat{h}_κ without specifying the range $\kappa = 0, 1, 2$, whenever statements are valid for all three cases. Further, for any function $\psi_h(x)$, we denote by $\psi^{[i]}$ the i -th derivative with respect to the “bandwidth” h ; and by ψ', ψ'' and $\psi^{(j)}$ the first, second and j -th derivative, respectively, with respect to x .

Proof of Theorem 3.1. The outline of the proof follows standard arguments, as described, for example, in Jones, Marron and Park (1991). Let \hat{h}_κ denote the minimizer of \hat{M}_κ . The proof is based on the representation

$$\hat{h}_\kappa - h_0 = \frac{\Delta_\kappa^{[1]}(\hat{h}_\kappa)}{MISE^{[2]}(h^*)}, \quad (5.1)$$

where $\Delta_\kappa(h) = [\hat{M}_\kappa - MISE_f](h)$, and h^* is some convex combination of \hat{h}_κ and h_0 . Under $\hat{h}/h_0 \xrightarrow{P} 1$, the denominator yields

$$MISE^{[2]}(h^*) = MISE^{[2]}(h_0) + o_p(n^{-2/5}).$$

Asymptotically, the randomness of \hat{h}_κ is determined by the numerator.

Define $v_m = D_h^{[1]} * D_h * L_{m,g}$, $w_m = K_h^{[1]} * K_h * L_{m,g}$, and

$$V_m = \frac{2}{n(n-1)} \sum_{i < j} v_m(X_i - X_j), \quad (5.2)$$

$$W_m = \frac{2}{n(n-1)} \sum_{i < j} w_m(X_i - X_j),$$

where $L_{0,g} = K_0$, $L_{1,g} = K_g$, and $L_{m,g} = L_{m-1,g} * K_g$ for $m = 1, \dots, 8$. Again, K_0 is the Dirac-operator with atom weight on zero. Note that the functions v_m and w_m as well as the corresponding U -statistics V_m and W_m depend on both bandwidths h and g .

For each criterion \hat{M}_κ , the numerator $\Delta_\kappa^{[1]}(\hat{h}_\kappa)$ can be represented as linear combinations of the U -statistics V_m and W_m , plus some remainder term of lower order, as shown in Lemma 5.1 below. Under the assumptions (A1) – (A3) the U -statistics V_m and W_m are asymptotically normally distributed, and so are the $(\hat{h}_\kappa - h_0)/h_0$. For details on U -statistics see Billingsley (1986). Lemma 5.2 provides expectations and covariances of the V_m and W_m .

Lemma 5.1 *Let $C_0 = R(D_h * f)^{[1]} - n^{-1}R(K_h * f)^{[1]}$. Then,*

$$\Delta_0^{[1]}(h) = 2V_2 - 2n^{-1}W_2 - C_0, \quad (5.3)$$

$$\Delta_1^{[1]}(h) = 4V_2 - 2V_4 - n^{-1}(4W_2 + 2W_4) - C_0, \quad (5.4)$$

$$\Delta_2^{[1]}(h) = 6V_2 - 6V_4 + 2V_6 - n^{-1}(6W_2 - 6W_4 + 2W_6) - C_0. \quad (5.5)$$

Put $\tilde{K}(x) = -K(x) - xK(x)$.

Lemma 5.2 *Assume that (A1) – (A3) hold, and that $h/g \rightarrow 0$, $g \rightarrow 0$ and $ng^{3/2} \rightarrow \infty$. Then, for all $m, m_1, m_2 = 1, \dots, 8$,*

$$\begin{aligned} \text{(i)} \quad \mathbf{E}V_m &= \frac{1}{2}R(D_h * f)^{[1]} + \sum_{i=1}^3 \sum_{l=2}^{5-i} g^{2i} h^{2l-1} \frac{(-1)^{i+l}}{(2i)!(2l)!} \mu_{2i}(L_m) \mu_{2l}(\tilde{K} * D) R(f^{(i+l)}) \\ &\quad + O(h^3 g^8) + o(h^9), \end{aligned} \quad (5.6)$$

$$\begin{aligned} \mathbf{E}W_m &= \frac{1}{2}R(K_h * f)^{[1]} + \sum_{i=1}^3 \sum_{l=1}^2 g^{2i} h^{2l-1} \frac{(-1)^{i+l}}{(2i)!(2l)!} \mu_{2i}(L_m) \mu_{2l}(\tilde{K} * K) R(f^{(i+l)}) \\ &\quad + O(hg^8) + o(h^5). \end{aligned} \quad (5.7)$$

$$\begin{aligned}
(ii) \quad \text{cov}(V_{m_1}, V_{m_2}) &= 4n^{-1} \sum_{l=2}^3 \sum_{k=2}^3 h^{2(l+k)-2} \frac{1}{(2l)!(2k)!} \mu_{2l}(\tilde{K} * D) \mu_{2k}(\tilde{K} * D) \\
&\quad \times \left\{ \sum_{i=0}^3 g^{2i} \Gamma_{i|k,l}(m_1, m_2) \right\} \\
&\quad + 2n^{-2} h^6 g^{-9} \frac{1}{(4!)^2} \mu_4(\tilde{K} * D)^2 R(f) \int L_{m_1}^{(4)} L_{m_2}^{(4)} \\
&\quad + o(n^{-1} h^6 g^6) + o(n^{-1} h^{10}) + o(n^{-2} h^6 g^{-9}), \tag{5.8}
\end{aligned}$$

where

$$\begin{aligned}
\Gamma_{i|k,l}(m_1, m_2) &= \sum_{j=0}^i \frac{1}{(2j)!(2i-2j)!} \mu_{2j}(L_{m_1}) \mu_{2i-2j}(L_{m_2}) \\
&\quad \times \left[\int f^{(2l+2j)} f^{(2k+2i-2j)} f - (-1)^{i+l+k} R(f^{(l+j)}) R(f^{(k+i-j)}) \right]; \tag{5.9}
\end{aligned}$$

$$\text{cov}(V_{m_1}, W_{m_2}) = O(n^{-1} h^4) + O(n^{-2} h^4 g^{-7}); \tag{5.10}$$

$$\text{cov}(W_{m_1}, W_{m_2}) = O(n^{-1} h^2) + O(n^{-2} h^2 g^{-5}). \tag{5.11}$$

The proofs of the lemmas are deferred to the end of this section.

Note that for all $m = 1, \dots, 8$,

$$\begin{aligned}
\mu_0(L_m) &= 1, & \mu_2(L_m) &= m\mu_2(K), \\
\mu_4(L_2) &= 2\mu_4(K) + 6\mu_2(K)^2, \\
\mu_4(L_m) &= m\mu_4(K) + 3m(m-1)\mu_2(K)^2, \\
\mu_6(L_2) &= 2\mu_6(K) + 30\mu_2(K)\mu_4(K), \\
\mu_6(L_m) &= m\mu_6(K) + 15m(m-1)\mu_2(K)\mu_4(K) + 15m(m-1)(m-2)\mu_2(K)^3, \tag{5.12} \\
\mu_0(\tilde{K} * D) &= \mu_2(\tilde{K} * D) = 0, & \mu_4(\tilde{K} * D) &= 12\mu_2(K)^2, \\
\mu_6(\tilde{K} * D) &= 90\mu_2(K)\mu_4(K), \\
\mu_0(\tilde{K} * L_m) &= 0, & \mu_2(\tilde{K} * L_m) &= 2\mu_2(K), \\
\mu_4(\tilde{K} * L_m) &= 4\mu_4(K) + 12m\mu_2(K)^2.
\end{aligned}$$

In order to derive $\mathbf{E}\Delta_\kappa^{[1]}$ and $\text{var}\Delta_\kappa^{[1]}$, use the representation in Lemma 5.1. According to Lemma 5.2 (i), the $\mathbf{E}V_m$ and $\mathbf{E}W_m$ depend on m only through the moments $\mu_{2i}(L_m)$ (up to some remainder terms). The moments yield

$$4\mu_2(L_2) - 2\mu_2(L_4) = 0,$$

$$6\mu_2(L_2) - 6\mu_2(L_4) + 2\mu_2(L_6) = 6\mu_4(L_2) - 6\mu_4(L_4) + 2\mu_4(L_6) = 0.$$

Hence, for fixed $i \leq \kappa$, the $g^{2i}h^{2\kappa-1}$ -terms of the double sums in (5.6) and in (5.7) cancel each other out in formulae (5.4) and (5.5), and we obtain:

$$\mathbf{E}\Delta_\kappa^{[1]}(h) = h^3 g^{2+2\kappa} \tilde{C}_{\mu,\kappa} + O(h^3 g^{4+2\kappa}) + o(h^9) + o(n^{-1}h^5), \quad (5.13)$$

$$\text{with } \tilde{C}_{\mu,\kappa} = -\mu_2(K)^{3+\kappa} R(f^{(3+\kappa)}).$$

The variance of $\Delta_\kappa^{[1]}(h)$ is dominated by the V_m -terms in formulae (5.3) - (5.5). In order to consolidate remainder terms, assume that h is of optimal order, $h \sim n^{-1/5}$. Using the results (5.8) - (5.11), we obtain from (5.3) - (5.5):

$$\text{var}\Delta_\kappa^{[1]}(h) = n^{-2}h^6 g^{-9} \tilde{C}_{1,\sigma} + n^{-1}h^6 \tilde{C}_{2,\sigma} + o(n^{-2}h^6 g^{-9}) + o(n^{-1}h^{10}) + O(n^{-1}h^6 g^{2+2\kappa}), \quad (5.14)$$

$$\text{where } \tilde{C}_{1,\sigma} = 2\mu_2(K)^4 R(f) R(L_2^{(4)}),$$

$$\tilde{C}_{2,\sigma} = 4\mu_2(K)^4 \left[\int (f^{(4)})^2 f - R(f'')^2 \right].$$

Note that for fixed k, l and $1 \leq i \leq \kappa$, the $\Gamma_{i|k,l}(m_1, m_2)$ terms in (5.8) cancel each other out when computing $\text{var}\Delta_\kappa^{[1]}$. Hence, we observe a remainder term of order $O(n^{-1}h^6 g^{2+2\kappa})$ rather than $O(n^{-1}h^6 g^2)$.

The assumptions (A1) - (A3) guarantee that $\hat{h}_\kappa/h_0 \xrightarrow{P} 1$; the argument is similar to Hall (1990). Hence, we can apply (5), and replace the denominator in formula (5.1) by

$$\text{MISE}^{[2]}(h_0) = n^{-2/5} C_D + o(n^{-2/5}),$$

with $C_D = 5R(f'')^{3/5} R(K)^{2/5} \mu_2(K)^{6/5}$.

Recall that \hat{h}_κ is defined to minimize the extrapolation criterion $\hat{M}_\kappa(h)$. For this case, the statement of Theorem 3.1 (i) then follows with

$$C_{\mu,\kappa,\eta} = \tilde{C}_{\mu,\kappa} C_D^{-1},$$

$$C_{1,\sigma} = \tilde{C}_{1,\sigma} C_D^{-2}, \quad \text{and} \quad C_{2,\sigma} = \tilde{C}_{2,\sigma} C_D^{-2}.$$

Similarly, Theorem 3.1 (ii) follows from (5.13) and (5.14) with $\hat{h}_\kappa/h_0 \rightarrow 1$; the given bandwidth g_κ is obtained as a solution to

$$\begin{aligned} \frac{\partial}{\partial g} \left\{ \mathbf{E}[\hat{h}_\kappa - h_0]^2 \right\} &= (4 + 4\kappa) h^6 g^{3+4\kappa} \tilde{C}_{\mu,\kappa}^2 - 9n^{-2} h^6 g^{-10} \tilde{C}_{1,\sigma} \\ &\quad + O(h^6 g^{5+4\kappa}) + O(n^{-1} h^6 g^{1+2\kappa}) + o(n^{-2} h^6 g^{-10}) \\ &= 0 \end{aligned}$$

for $h = h_0 = n^{-1/5} R(f'')^{-1/5} R(K)^{1/5} \mu_2(K)^{-2/5} + o(n^{-1/5})$.

In order to prove Theorem 3.1 for the normal reference criterion $\hat{N}_\kappa(h)$, we have to replace the numerator in (5.1) by

$$\tilde{\Delta}_\kappa^{[1]}(h) = \Delta_\kappa^{[1]}(h) - \left\{ \mathbf{E} \hat{M}_\kappa(h, \hat{\eta}_g) - MISE_\eta \right\}^{[1]}, \quad (5.15)$$

while \hat{h}_κ now denotes the minimizer of $\hat{N}_\kappa(h)$.

The additional term $\{\mathbf{E} \hat{M}_\kappa(h, \hat{\eta}_g) - MISE_\eta\}^{[1]}$ is not random and thus affects only the bias of $(\hat{h}_\kappa - h_0)$. Note that $\hat{M}_\kappa(h; \hat{\eta}_g)$ is simply \hat{M}_κ applied to a sample drawn from η (instead of f). Hence, formulae (5.13) and (5.15) provide

$$\mathbf{E} \tilde{\Delta}_\kappa^{[1]}(h) = h^3 g^{2+2\kappa} \tilde{C}_{\mu, \kappa, \eta} + O(h^3 g^{4+2\kappa}) + o(h^{10})$$

$$\text{with } \tilde{C}_{\mu, \kappa, \eta} = -\mu_2(K)^{3+\kappa} [R(f^{(3+\kappa)}) - R(\eta^{(3+\kappa)})].$$

The variance and the convergence rate of the bias of $(\hat{h}_\kappa - h_0)$ are unchanged, and with the new constant $C_{\mu, \kappa, \eta}$, the statements of Theorem 3.1 follow as before. \square

Proof of Lemma 5.1. The criteria \hat{M}_κ are linear combinations of quantities $\widetilde{MISE}_{L_{m,g} * \hat{f}_g}$, $m = 0, 1, 2$, as defined in formulae (2.5) - (2.7). Note that the $L_{m,g} * \hat{f}_g$ are simply kernel estimates with bandwidth g and kernel L_{m+1} . Hence, $\widetilde{MISE}_{L_{m,g} * \hat{f}_g}$ is defined through formula (2.4) as:

$$\begin{aligned} \widetilde{MISE}_{L_{m,g} * \hat{f}_g}(h) &= \frac{n}{n-1} R(D_h * L_{m,g} * \hat{f}_g) - \frac{1}{n-1} R(D_h * L_{m+1,g}) + \frac{1}{n} R(K_h) \\ &\quad - \frac{1}{n-1} R(K_h * L_{m,g} * \hat{f}_g) + \frac{1}{n(n-1)} R(K_h * L_{m+1,g}). \end{aligned} \quad (5.16)$$

In the first part of the proof we provide some useful representations of terms in (5.16); the conjectures of Lemma 5.1 then follow through straightforward calculations.

Note that $\psi^{[1]}$ denotes the first derivative with respect to h . Corresponding derivatives of terms in (5.16) can be written as follows:

$$\begin{aligned} R(D_h * L_{m,g} * \hat{f}_g)^{[1]} &= n^{-2} \sum_{i=1}^n \sum_{j=1}^n 2D_h^{[1]} * D_h * L_{2m+2}(X_i - X_j) \\ &= 2n^{-1} D_h^{[1]} * D_h * L_{2m+2}(0) + 4n^{-2} \sum_{i < j} D_h^{[1]} * D_h * L_{2m+2}(X_i - X_j) \\ &= 2n^{-1} v_{2m+2}(0) + 2 \frac{n-1}{n} V_{2m+2}; \end{aligned} \quad (5.17)$$

$$\begin{aligned}
R(K_h * L_{m,g} * \hat{f}_g)^{[1]} &= n^{-2} \sum_{i=1}^n \sum_{j=1}^n 2K_h^{[1]} * K_h * L_{2m+2}(X_i - X_j) \\
&= 2n^{-1} w_{2m+2}(0) + 2 \frac{n-1}{n} W_{2m+2}; \tag{5.18}
\end{aligned}$$

$$R(D_h * L_{m,g})^{[1]} = 2v_{2m}(0); \tag{5.19}$$

$$R(K_h * L_{m,g})^{[1]} = 2w_{2m}(0). \tag{5.20}$$

Inserting formulae (5.17) - (5.20) in (5.16) results in:

$$\widetilde{MISE}_{L_{m,g} * \hat{f}_g}^{[1]}(h) = 2V_{2m+2} - 2n^{-1}W_{2m+2} + n^{-1}R(K_h)^{[1]}.$$

The assertion follows immediately with (2.5) - (2.7), and with

$$MISE_f(h) = R(D_h * f) - n^{-1}R(K_h * f) + n^{-1}R(K_h).$$

□

The proof of Lemma 5.2 is based on techniques used in Jones, Marron and Park (1991), relying heavily on Taylor expansions. However, some of the technical details are different (for example, we need deeper expansions), so that we decided to show the calculations for Lemma 5.2 in some detail.

Now and below, define $\psi_h(x) = h^{-1}\psi(x/h)$ for any function ψ and $h > 0$. The following Lemma provides asymptotic expansions that are useful for proving Lemma 5.2.

Lemma 5.3 *Assume that X_1, X_2 and X_3 are independent random variables with marginal density f , and that f satisfies (A3). Let a, b, α , and β denote symmetric functions, centered around 0, with $\mu_{10}(a), \mu_{10}(\alpha) < \infty$, $\mu_6(b), \mu_6(\beta) < \infty$; $\mu_i(a) = 0$ for $i = 0, \dots, (2r-1)$, and $\mu_i(\alpha) = 0$ for $i = 0, \dots, (2s-1)$.*

Then, for $h, g \rightarrow 0$, and $h/g \rightarrow 0$,

$$(i) \quad \mathbf{E}[a_h * b_g](X_1 - X_2) = \sum_{i=0}^3 \sum_{l=r}^{5-i} g^{2i} h^{2l} \frac{(-1)^{i+l}}{(2i)!(2l)!} \mu_{2i}(b) \mu_{2l}(a) R(f^{(i+l)}) + O(h^{2r} g^8) + o(h^{10}).$$

(ii) *Assume that b and β are r and s times differentiable, respectively, with Hölder-continuous derivatives. Then,*

$$\begin{aligned}
&\mathbf{E}[a_h * b_g](X_1 - X_2) [\alpha_h * \beta_g](X_1 - X_2) \\
&= h^{2(r+s)} g^{-2(r+s)-1} \frac{1}{(2r)!(2s)!} \mu_{2r}(a) \mu_{2s}(\alpha) R(f) \int b^{(2r)} \beta^{(2s)} + o(h^{2(r+s)} g^{-2(r+s)-1}) \tag{5.21}
\end{aligned}$$

and

$$\begin{aligned} & \mathbf{E} [a_h * b_g] (X_1 - X_2) [\alpha_h * \beta_g] (X_1 - X_3) \\ &= \sum_{l=r}^3 \sum_{k=s}^3 h^{2(l+k)} \frac{1}{(2l)!(2k)!} \mu_{2l}(a) \mu_{2k}(\alpha) \left\{ \sum_{i=0}^3 g^{2i} \tilde{\Gamma}_{i|k,l} \right\} + o(h^{2r+2s}g^6) + o(h^{12}), \end{aligned} \quad (5.22)$$

where

$$\tilde{\Gamma}_{i|k,l} = \sum_{j=0}^i \frac{1}{(2j)!(2i-2j)!} \mu_{2j}(b) \mu_{2i-2j}(\beta) \int f^{(2l+2j)} f^{(2k+2i-2j)} f. \quad (5.23)$$

The proof is based on suitable variable substitution and subsequent Taylor expansions of f [and of b and β in (5.21)] in the integrals. The technique is described in Jones, Marron and Park (1991, p. 1929-31), where similar statements are proven. The assumptions on f insure the necessary uniform convergence of remainder terms. We omit the proof of Lemma 5.3.

Proof of Lemma 5.2. The function $\tilde{K}(x) = -K(x) - xK'(x)$ is defined such that $D_h^{[1]}(x) = K_h^{[1]}(x) = h^{-1}\tilde{K}_h(x)$, and, consequently,

$$v_m(X_i - X_j) = h^{-1}\tilde{K}_h * D_h * L_{m,g}(X_i - X_j). \quad (5.24)$$

Set $a = \tilde{K} * D$, and $b = L_m$. Then, a and b satisfy the conditions of Lemma 5.3, with $\mu_i(a) = 0$ for $i \leq 3$; moments of a and b (with respect to f) are given in (5.12). In order to derive $\mathbf{E}V_m$, we apply Lemma 5.3 (i) to (5.24), resulting in:

$$\begin{aligned} \mathbf{E}V_m &= \mathbf{E}v_m(X_1 - X_2) \\ &= \sum_{i=0}^3 \sum_{l=2}^{5-i} g^{2i} h^{2l-1} \frac{(-1)^{i+l}}{(2i)!(2l)!} \mu_{2i}(L_m) \mu_{2l}(\tilde{K} * D) R(f^{(i+l)}) \\ &\quad + O(h^3 g^8) + o(h^9). \end{aligned} \quad (5.25)$$

Note that the terms in the double sum in (5.25) are the same as in (5.6), with the only difference that the sum in (5.25) starts at $i = 0$. The additional h^{2l-1} -terms ($i = 0$) are approximated by $\frac{1}{2}R(D_h * f)^{[1]}$; using Taylor expansions of f around x , we obtain:

$$\begin{aligned} \frac{1}{2}R(D_h * f)^{[1]} &= h^{-1} \int [\tilde{K}_h * f](x) [D_h * f](x) dx \\ &= h^{-1} \iiint \tilde{K}(t) D(s) f(x - th) f(x - sh) ds dt dx \end{aligned}$$

$$\begin{aligned}
&= h^{-1} \iint \tilde{K}(t) D(s) \sum_{k=0}^5 \frac{(-1)^k}{(2k)!} (t+k)^{2k} h^{2k} R(f^{(k)}) ds dt + o(h^9) \\
&= \sum_{k=2}^5 h^{2k-1} \mu_{2k}(\tilde{K} * D) \frac{(-1)^k}{(2k)!} R(f^{(k)}) + o(h^9). \tag{5.26}
\end{aligned}$$

Formula (5.6) follows immediately from (5.26) and (5.25). The expectation $\mathbf{E}W_m$ can be computed similarly, using $a = \tilde{K} * K$ in Lemma 5.3.

In order to compute $\text{cov}(V_{m_1}, V_{m_2})$, we use again representation (5.2). Let X_1, \dots, X_4 be independent random variables with density f . Then,

$$\begin{aligned}
\mathbf{E}V_{m_1}V_{m_2} &= \frac{2}{n(n-1)} \mathbf{E}v_{m_1}(X_1 - X_2)v_{m_2}(X_1 - X_2) \\
&\quad + \frac{4(n-2)}{n(n-1)} \mathbf{E}v_{m_1}(X_1 - X_2)v_{m_2}(X_1 - X_3) \\
&\quad + \frac{(n-2)(n-3)}{n(n-1)} \mathbf{E}v_{m_1}(X_1 - X_2)\mathbf{E}v_{m_2}(X_3 - X_4).
\end{aligned}$$

Consequently,

$$\begin{aligned}
\text{cov}(V_{m_1}, V_{m_2}) &= \left\{ -4n^{-1} \mathbf{E}V_{m_1} \mathbf{E}V_{m_2} + 4n^{-1} \mathbf{E}v_{m_1}(X_1 - X_2)v_{m_2}(X_1 - X_3) \right. \\
&\quad \left. + 2n^{-2} \mathbf{E}v_{m_1}(X_1 - X_2)v_{m_2}(X_1 - X_2) \right\} \left\{ 1 + O(n^{-1}) \right\}. \tag{5.27}
\end{aligned}$$

The expectations in (5.27) are again computed through Lemma 5.3. Set $a = \alpha = \tilde{K} * D$, $b = L_{m_1}$, $\beta = L_{m_2}$. Then, with (5.24) and (5.21), we obtain for the last term in (5.27):

$$\begin{aligned}
\mathbf{E}v_{m_1}(X_1 - X_2)v_{m_2}(X_1 - X_2) &= h^6 g^{-9} \frac{1}{(4!)^2} \mu_4(\tilde{K} * D)^2 R(f) \int L_{m_1}^{(4)} L_{m_2}^{(4)} \\
&\quad + o(h^6 g^{-9}). \tag{5.28}
\end{aligned}$$

For the second term in (5.27) we use (5.22):

$$\begin{aligned}
\mathbf{E}v_{m_1}(X_1 - X_2)v_{m_2}(X_1 - X_3) &= \sum_{l=2}^3 \sum_{k=2}^3 h^{2(l+k)-2} \frac{1}{(2l)!(2k)!} \mu_{2l}(\tilde{K} * D) \mu_{2k}(\tilde{K} * D) \\
&\quad \times \left\{ \sum_{i=0}^3 g^{2i} \tilde{\Gamma}_{i|k,l}(m_1, m_2) \right\} + o(h^6 g^6) + o(h^{10}), \tag{5.29}
\end{aligned}$$

where $\tilde{\Gamma}_{i|k,l}(m_1, m_2)$ is defined by (5.23) with $b = L_{m_1}$ and $\beta = L_{m_2}$. For the $\mathbf{E}V_{m_i}$ in the first term of (5.27) we use the representation (5.25). Formula (5.8) then follows from (5.27) by inserting (5.28), (5.29) and (5.25).

Formulae (5.10) and (5.11) are derived similarly, with appropriate choice of a and α . \square

References

- [1] Becker, R.A., Chambers, J.M. and Wilks, A.R. (1988), *The New S Language*, Wadsworth & Brooks/COLE Computer science series.
- [2] Billingsley, P. (1986), *Probability and Measure*, John Wiley, New York.
- [3] Bowman, A.W. (1984), An alternative method of cross-validation for the smoothing of density estimates, *Biometrika*, 71, 353-360.
- [4] Faraway, J.J. and Jhun, M. (1990), Bootstrap choice of bandwidth for density estimation, *J. Amer. Statist. Assoc.*, 85, 1119 - 1122.
- [5] Grund, B., Hall, P. and Marron, J.S. (1994), On loss and risk in smoothing parameter selection. *J. Nonparametric Statist.*, 4, 107-32.
- [6] Hall, P. (1990), Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems, *Journal of Multivariate Analysis*, 32, 177-203.
- [7] Hall, P. and Marron, J.S. (1987a), Estimation of integrated squared density derivatives, *Statistics & Probability Letters*, 6, 109-115.
- [8] Hall, P. and Marron, J.S. (1987b), Extent to which least-squares cross-validation minimises integrated square error in nonparametric density estimation. *Probab. Th. Rel. Fields*, 74, 567-581.
- [9] Hall, P. and Marron, J.S. (1991), Lower bounds for bandwidth selection in density estimation. *Probab. Th. Rel. Fields*, 90, 149-173.
- [10] Hall, P., Marron, J.S. and Park, B. U. (1992), Smoothed cross-validation, *Probab. Th. Rel. Fields*, 92, 1-20.
- [11] Jones, M.C. (1991), The roles of ISE and MISE in density estimation, *Statistics & Probability Letters*, 12, 51-56.
- [12] Jones, M.C. and Kappenman, R.F. (1991), On a class of kernel density estimate bandwidth selectors, *Scand. J. Statist.*, 19, 337-349.
- [13] Jones, M.C., Marron, J.S. and Park, B.U. (1991), A simple root n bandwidth selector, *The Annals of Statistics*, 19, 1919-1932.
- [14] Jones, M.C., Marron, J.S. and Sheater, S.J.(1992), Progress in data-based bandwidth selection for kernel density estimation, unpublished manuscript.

- [15] Marron, J.S. and Wand, M.P. (1992), Exact mean integrated squared error, *The Annals of Statistics*, 20, 712-736.
- [16] Park, B.U. and Marron, J.S. (1990), Comparison of data-driven bandwidth selectors, *J. Amer. Statist. Assoc.*, 85, 66-72.
- [17] Park, B.U. and Turlach, B.A. (1992), Practical performance of several data driven bandwidth selectors, *Computational Statistics*, 7, 251-270.
- [18] Rudemo, M. (1982), Empirical choice of histograms and kernel density estimators, *Scandinavian Journal of Statistics*, 9, 65-78.
- [19] Scott, D. W. (1992), *Multivariate Density Estimation*, John Wiley, New York.
- [20] Sheater, S.J. and Jones, M.C. (1991), A reliable data-based bandwidth selection method for kernel density estimation, *J. R. Statist. Soc. B*, 53, 683-690.
- [21] Silverman, B.W. (1986), *Density Estimation*, Chapman and Hall, London, New York.
- [22] Silverman, B. W. and Young, G. A. (1987), The bootstrap: To smooth or not to smooth?, *Biometrika*, 74, 469-479.
- [23] Taylor, C.C. (1989), Bootstrap choice of the smoothing parameter in kernel density estimation, *Biometrika*, 76, 705-712.

Captions

Figure 1: Gaussian Density. Boxplots of $ISE(\hat{h})$ values for sample sizes $n = 100, 400$ and 1600 , obtained from 500 simulations. For \hat{h}_{LSCV} values larger than $1.5n^{-4/5}$ are set to $1.5n^{-4/5}$ (16, 6 and 1 value for sample sizes 100, 400 and 1600 respectively).

Figure 2: Gaussian Density: Boxplots of $(\ln \hat{h} - \ln \hat{h}_{ISE})$ values, obtained from 500 simulations for sample sizes 100, 400 and 1600.

Figure 3: Gaussian Density: Boxplots of $(\ln \hat{h} - \ln \hat{h}_{MISE})$ values, obtained from 500 simulations for sample sizes 100, 400 and 1600.

Figure 4: Strongly Skewed Density # 3 of Marron and Wand (1992). Boxplots of $ISE(\hat{h})$ values for sample sizes $n = 100, 400$ and 1600 , obtained from 500 simulations. Values \hat{h} larger than $6.6n^{-4/5}$ are set to $6.6n^{-4/5}$. ($n = 100$: one value for \hat{h}_{LSCV} ; $n = 400$: one value each for \hat{h}_{M_2} , \hat{h}_{N_2} , \hat{h}_{RN} and \hat{h}_{SJ} , 6 values for \hat{h}_{PM} ; $n = 1600$: 1 value each for \hat{h}_{SJ} and \hat{h}_{PM}).

Figure 5: Strongly Skewed Density # 3 of Marron and Wand (1992). Boxplots of $(\ln \hat{h} - \ln \hat{h}_{ISE})$ values for sample sizes 100, 400 and 1600, obtained from 500 simulations.

Figure 6: Claw Density # 10 of Marron and Wand (1992): Boxplots of $ISE(\hat{h})$ values for sample sizes $n = 100, 400$ and 1600 , obtained from 500 simulations. For $n = 1600$ values larger than $3.9n^{-4/5}$ are set to $3.9n^{-4/5}$ (7 values each for \hat{h}_{PM} and \hat{h}_{SJ}).

Figure 7: Claw Density # 10 of Marron and Wand (1992): Boxplots of $(\ln \hat{h} - \ln \hat{h}_{ISE})$ values obtained from 500 simulations for sample sizes 100, 400 and 1600.

Figure 8: Asymmetric Bimodal Density # 8 of Marron and Wand (1992). Boxplots of $ISE(\hat{h})$ values for sample sizes $n = 100, 400$ and 1600 , obtained from 500 simulations. Values larger than $1.6n^{-4/5}$ are set to $1.6n^{-4/5}$. ($n = 100$: 15 values for \hat{h}_{LSCV} and one value for \hat{h}_{MISE} ; $n = 400$: one value for \hat{h}_{LSCV} ; $n = 1600$: 4 values for \hat{h}_{LSCV} , one value each for \hat{h}_{M_2} , \hat{h}_{N_2} , \hat{h}_{RN} , \hat{h}_{SJ} , \hat{h}_{PM} and \hat{h}_{FJ}).

Figure 9: Asymmetric Bimodal Density # 8 of Marron and Wand (1992). Boxplots of $(\ln \hat{h} - \ln \hat{h}_{ISE})$ values for sample sizes $100, 400$ and 1600 , obtained from 500 simulations.

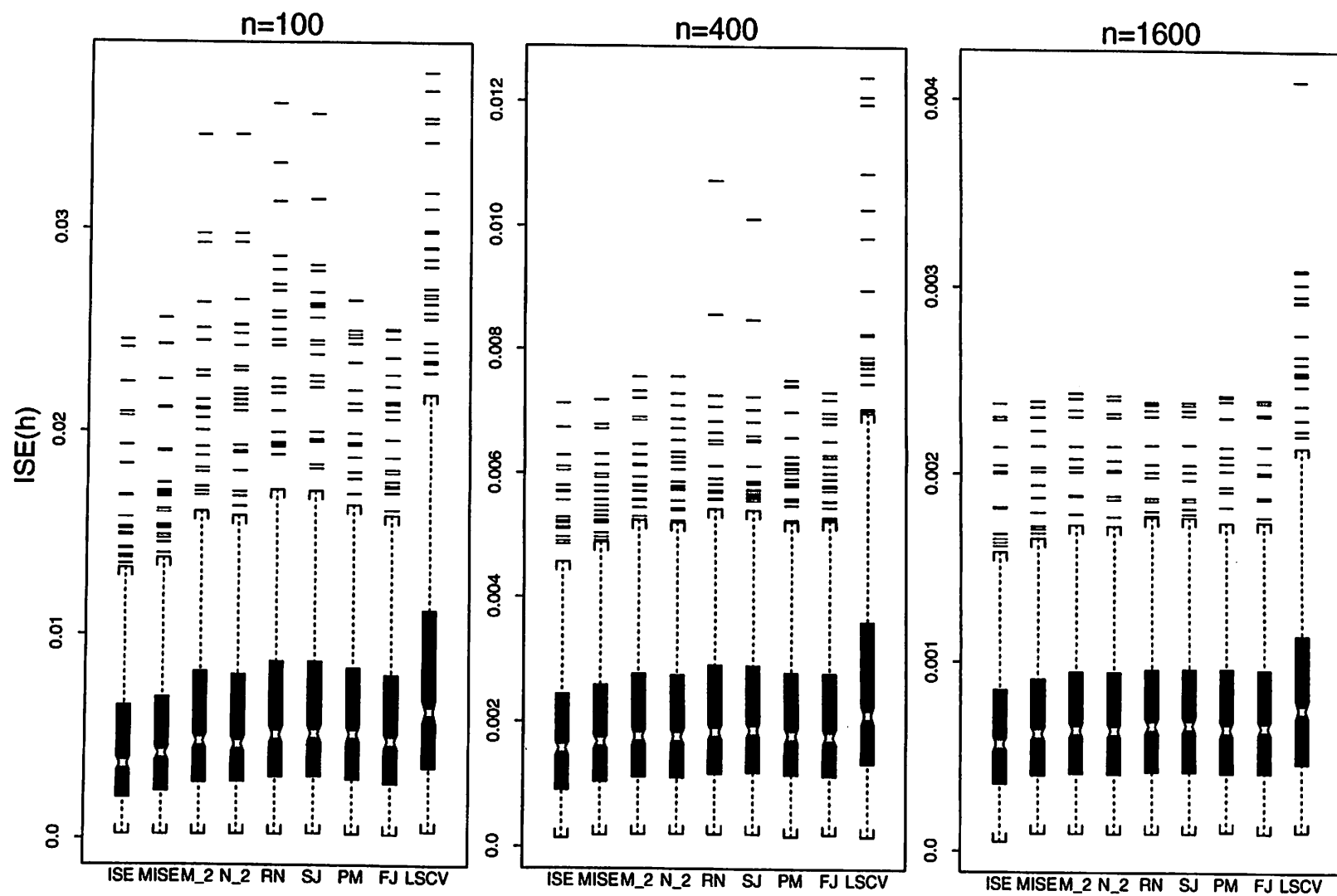


Figure 1

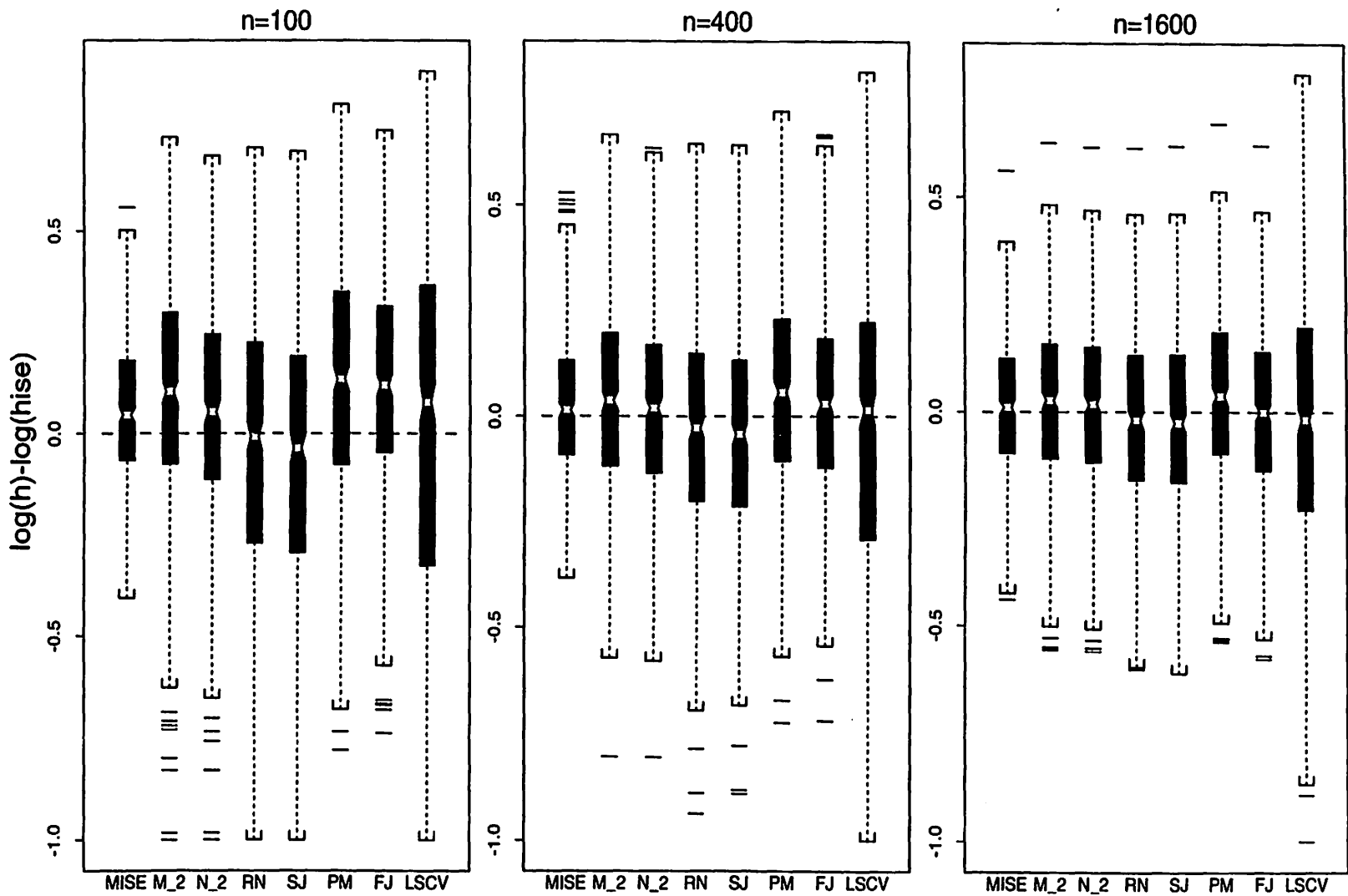


Figure 2

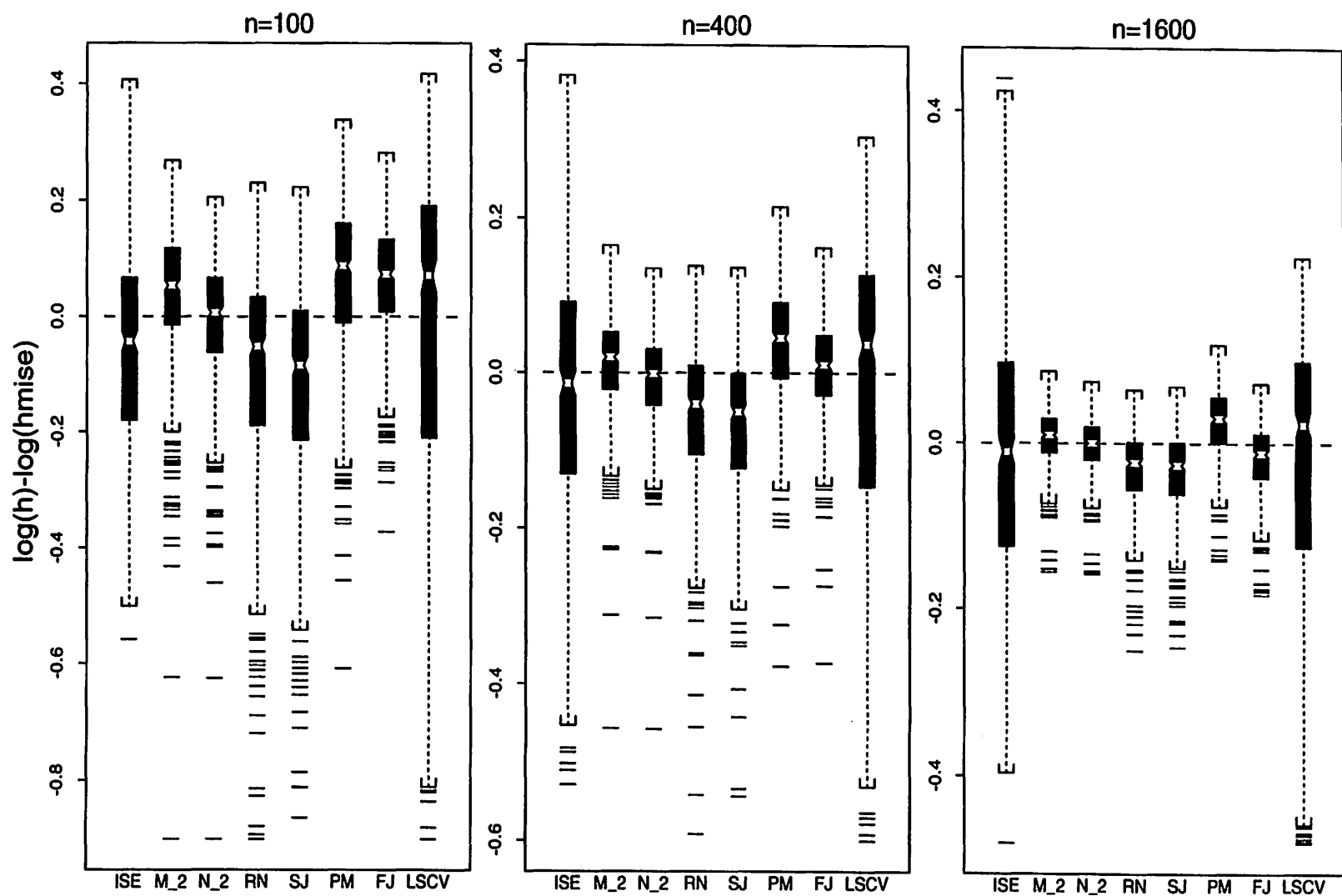


Figure 3

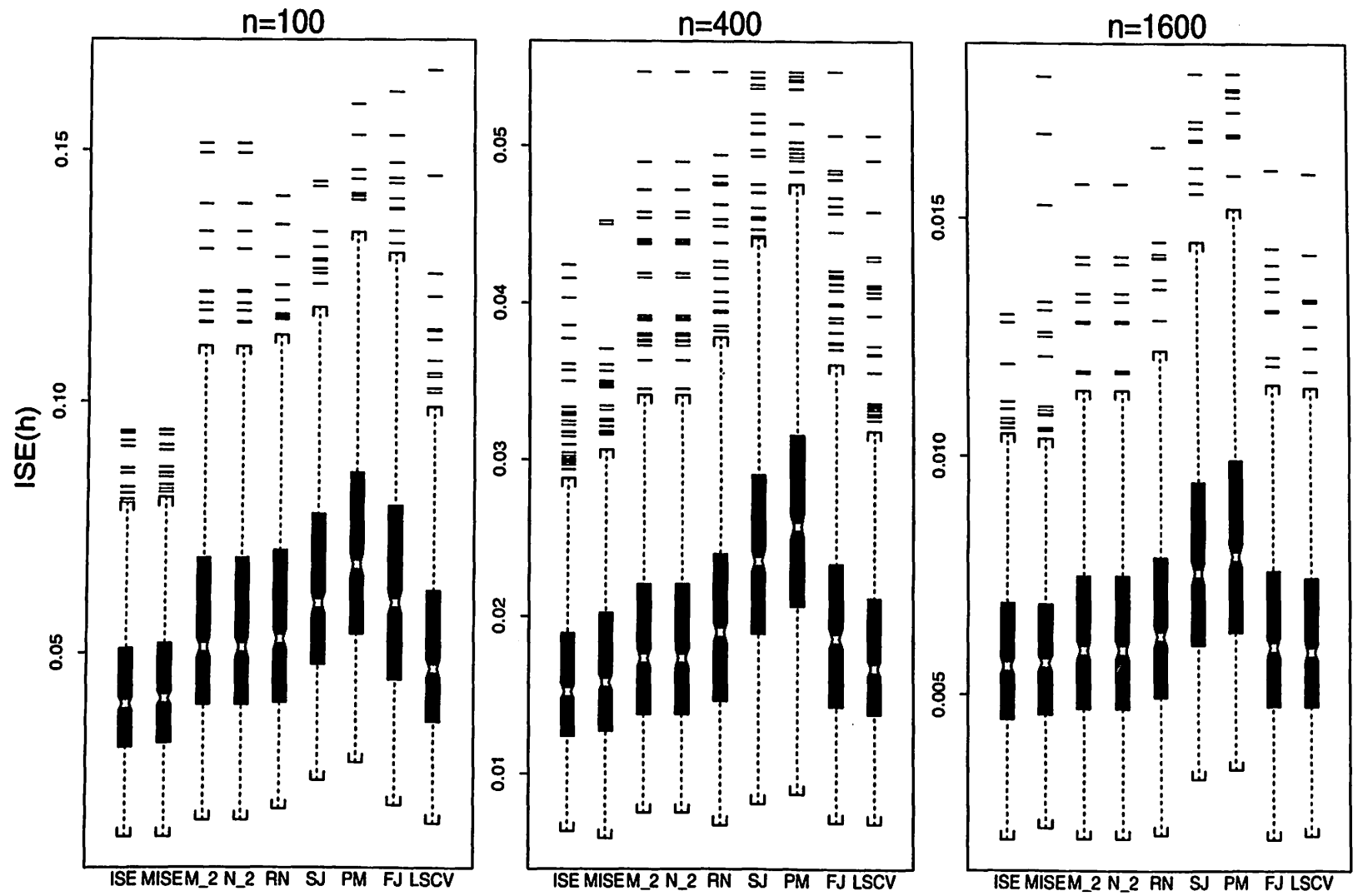


Figure 4

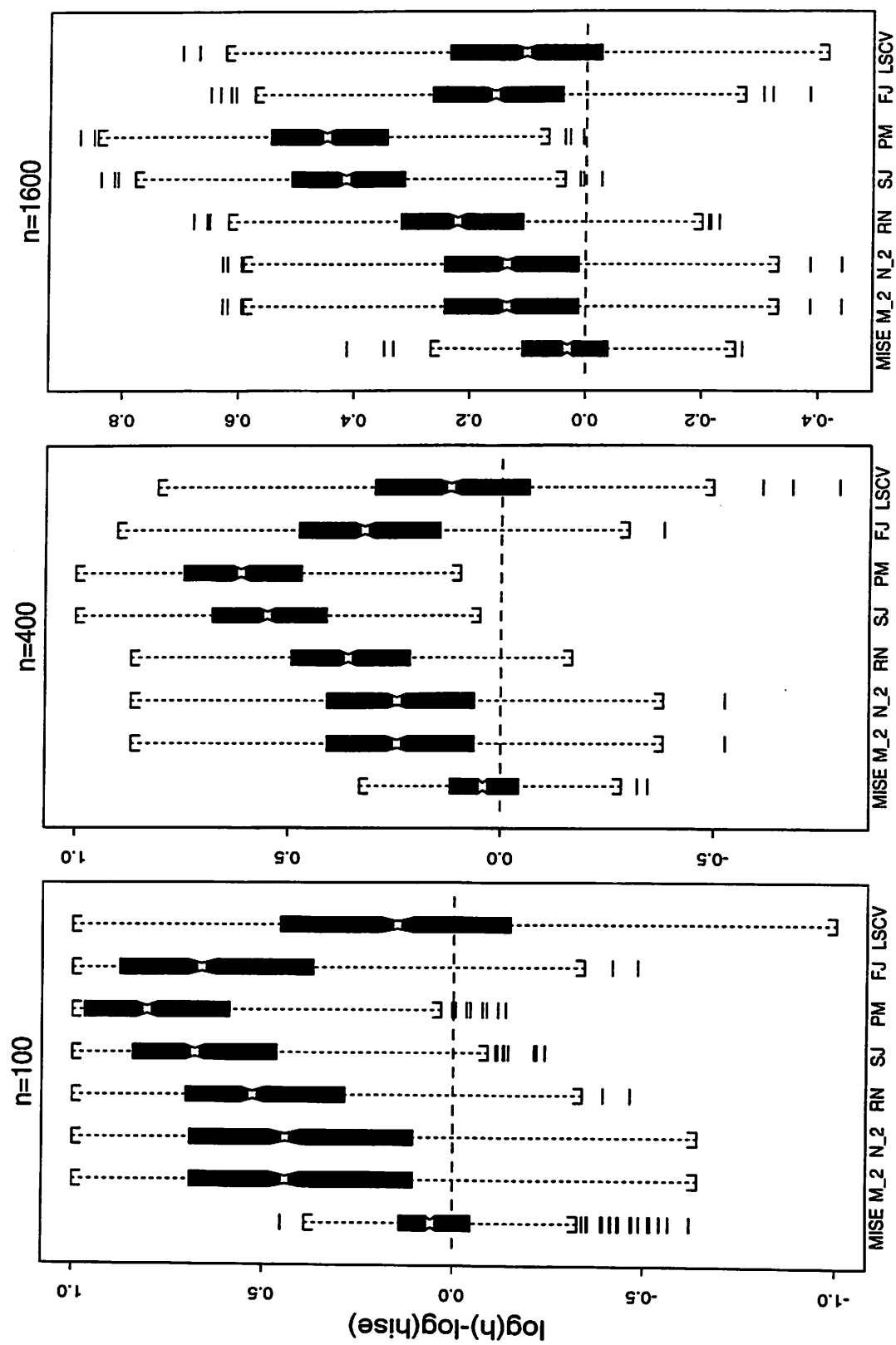


Figure 5

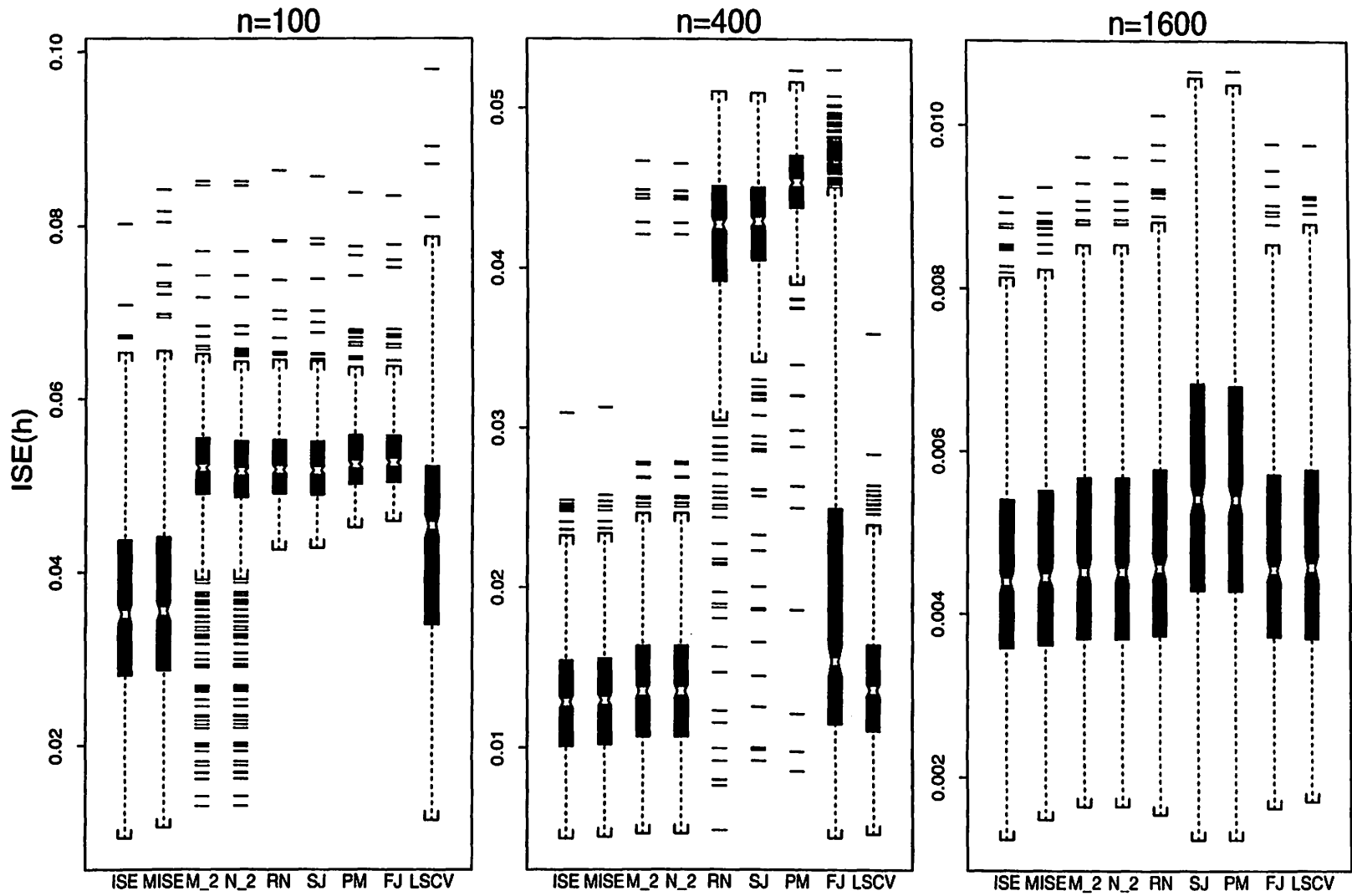


Figure 6

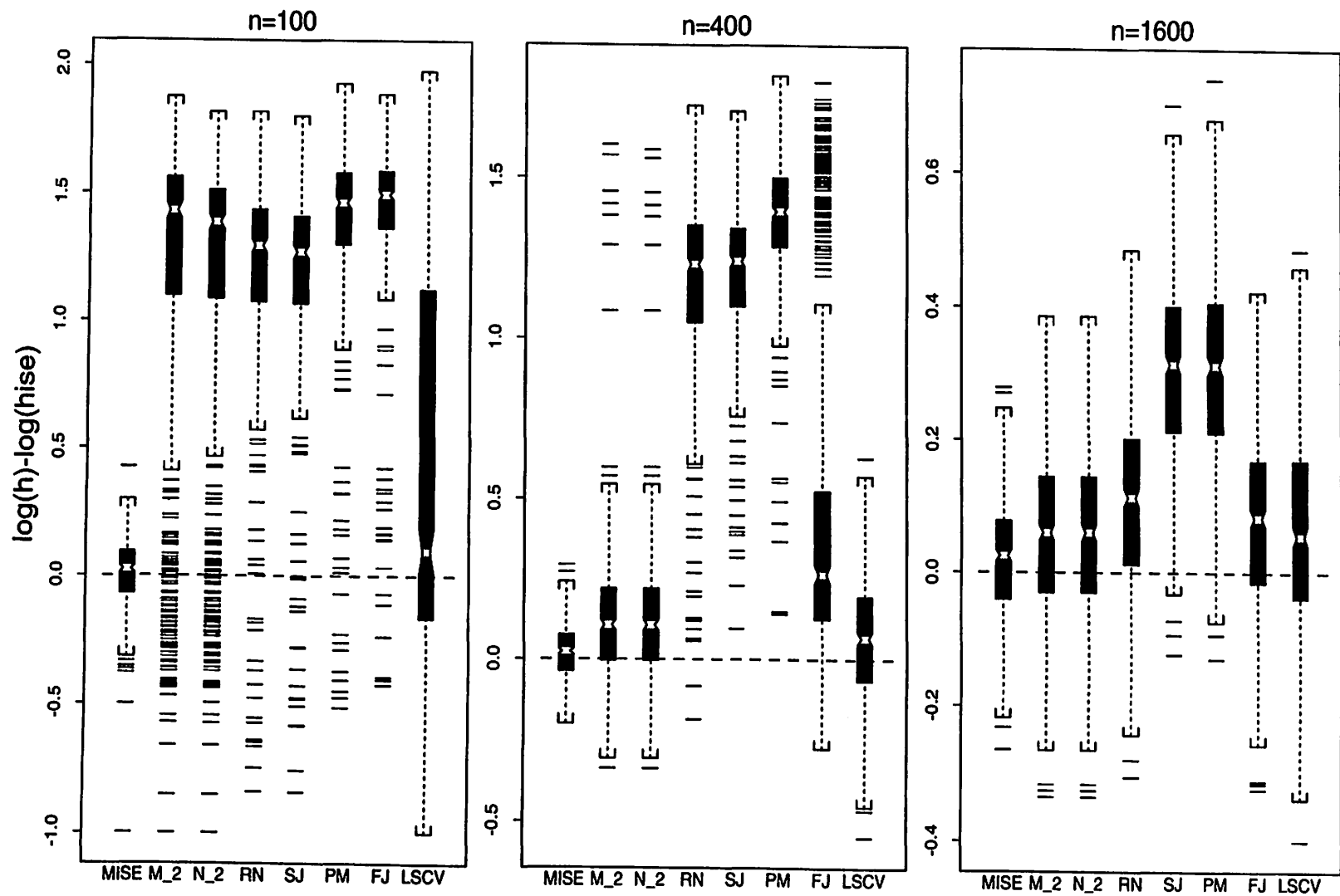


Figure 7

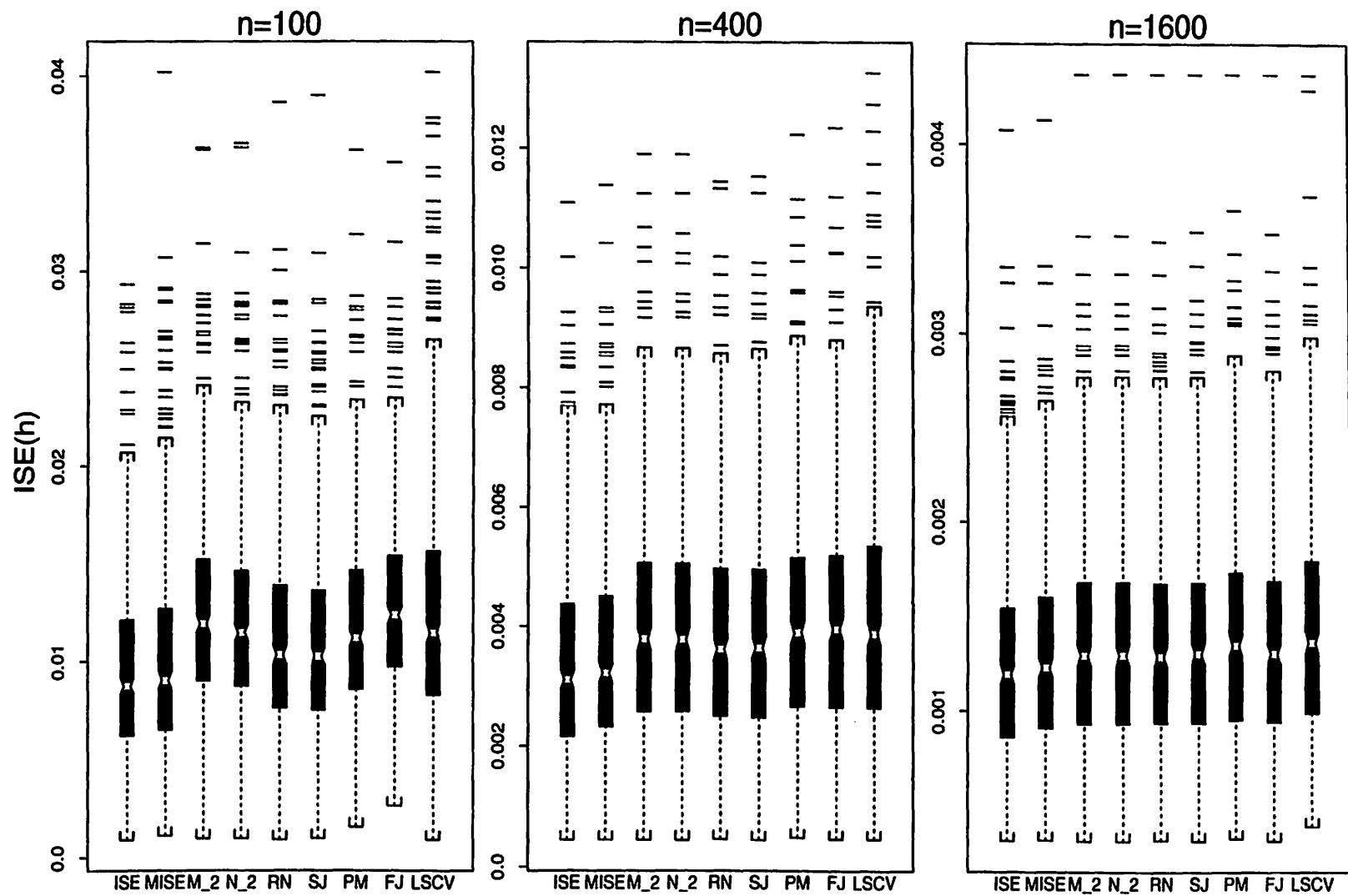


Figure 8

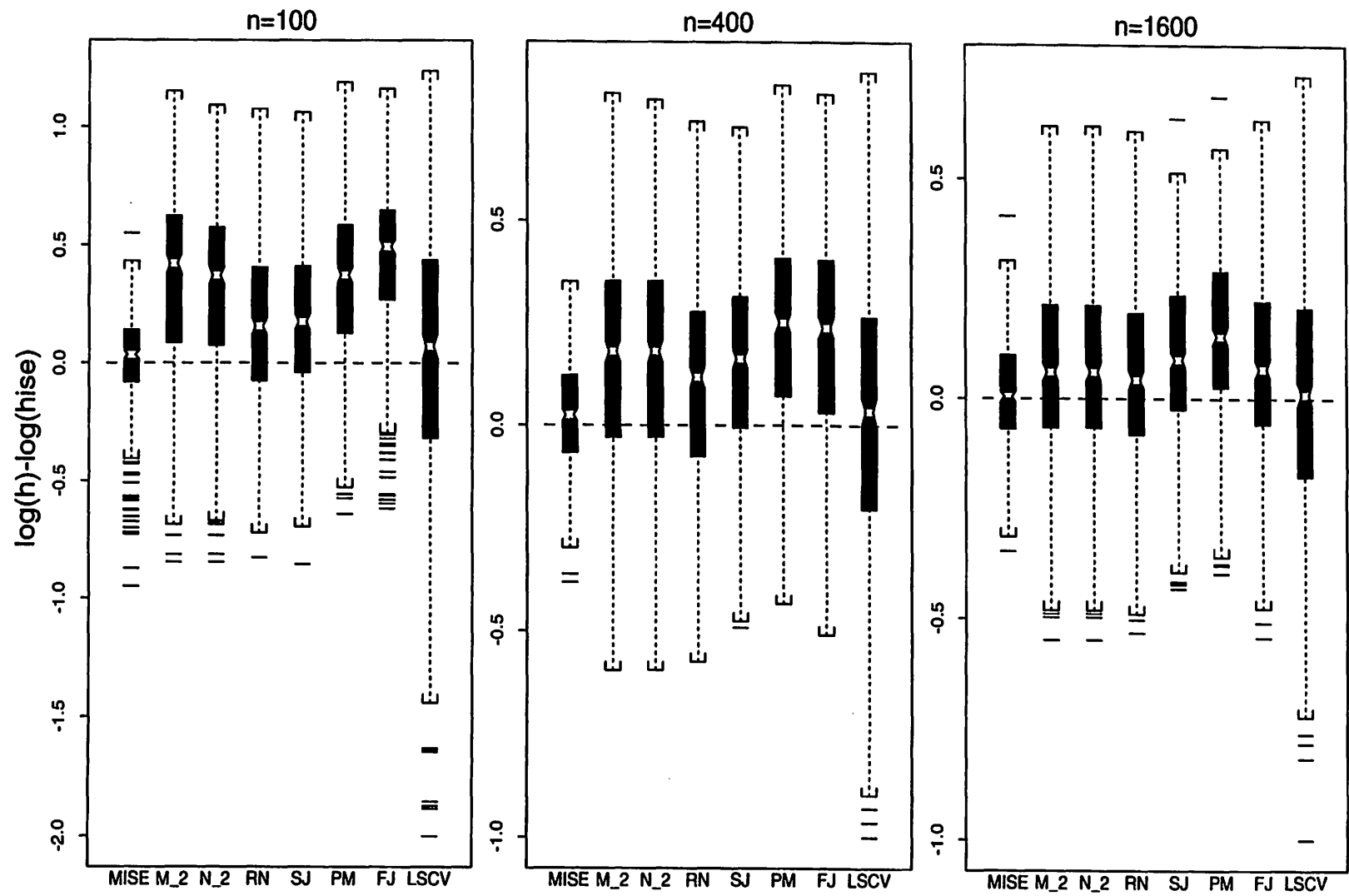


Figure 9

Table 1: Averages of $[(\hat{h}-h_0)/h_0]^2$ obtained for densities # 1 - 15 of Marron and Wand (1992). Sample sizes are $n = 100, 400$ and 1600 . v_{min} and v_{max} denote the minimum and maximum value of the coefficient of variation for the 12 bandwidth selectors.

d	n	ISE	\hat{M}_0	\hat{M}_1	\hat{M}_2	\hat{N}_0	\hat{N}_1	\hat{N}_2	RN	SJ	PM	LSCV	FJ	v_{min}	v_{max}
1	100	.0322	.0373	.02	.0185	.0084	.0096	.0137	.0313	.0332	.0256	.0975	.0164	.04	.12
1	400	.0293	.0103	.0051	.0047	.0027	.003	.0038	.0096	.011	.0077	.0608	.0043	.05	.11
1	1600	.0268	.0035	.0014	.0012	.0008	.0009	.001	.003	.0034	.0026	.0387	.0018	.03	.12
2	100	.0299	.0536	.0296	.0253	.0149	.0156	.0198	.0316	.0315	.0317	.0966	.0257	.05	.09
2	400	.0279	.0166	.0085	.0067	.0063	.0056	.0057	.0078	.0079	.0106	.0541	.0061	.04	.08
2	1600	.0266	.0055	.0025	.002	.0024	.0019	.0019	.003	.0029	.0039	.0371	.0022	.04	.1
3	100	.0333	1.3	.7883	.4339	1.279	.7864	.4338	.4412	.8098	1.333	.1617	.901	.03	.1
3	400	.015	.2109	.1399	.0896	.2108	.1399	.0896	.167	.4735	.6374	.0487	.1434	.02	.07
3	1600	.0122	.0443	.027	.0184	.0443	.027	.0184	.046	.2184	.2664	.0186	.0243	.01	.06
4	100	.0208	1.067	.5996	.2712	.8321	.5389	.2608	.132	.183	.5251	.1094	.7498	.08	.22
4	400	.0159	.0528	.0292	.0193	.0526	.0292	.0193	.016	.0396	.0888	.0266	.0312	.04	.07
4	1600	.0173	.0134	.0068	.005	.0134	.0068	.005	.0042	.0146	.0303	.0127	.0053	.03	.07
5	100	.0274	.0412	.0215	.0168	.0139	.0131	.0138	.0229	.0262	.0247	.0441	.0188	.05	.07
5	400	.0296	.0131	.0064	.0053	.0047	.0042	.0045	.0078	.0093	.0092	.0245	.0048	.05	.09
5	1600	.0322	.0045	.0019	.0015	.0017	.0013	.0013	.002	.0026	.0035	.0172	.0014	.04	.1
6	100	.0318	.3966	.2462	.1426	.1282	.127	.1056	.0517	.044	.1307	.1366	.2027	.03	.12
6	400	.0179	.0654	.0368	.0237	.0471	.0325	.0229	.0144	.0144	.0367	.0587	.0319	.05	.07
6	1600	.0189	.0116	.0052	.0036	.0105	.0051	.0036	.003	.0033	.0096	.0334	.0038	.04	.09
7	100	.0177	.0399	.0185	.0138	.0396	.0184	.0138	.0083	.0253	.0383	.0591	.0184	.04	.07
7	400	.0157	.0116	.0053	.0041	.0115	.0053	.0041	.0029	.0105	.0142	.0373	.0038	.04	.08
7	1600	.0163	.004	.0016	.0013	.004	.0016	.0013	.0011	.0043	.0059	.0265	.0013	.02	.09
8	100	.0578	.6185	.4175	.2707	.2554	.2505	.2098	.0916	.0828	.2356	.1912	.3531	.02	.15
8	400	.0184	.1673	.1039	.0594	.1295	.0945	.058	.0308	.0387	.0877	.0613	.082	.04	.07
8	1600	.0154	.0276	.0149	.0102	.0266	.0148	.0102	.0077	.0134	.029	.031	.0115	.04	.08
9	100	.0399	.5395	.3318	.1943	.2191	.1983	.1561	.0774	.0766	.1906	.1636	.2941	.03	.09
9	400	.0223	.1608	.103	.0685	.1393	.0984	.0676	.0387	.0563	.1036	.0666	.086	.04	.07
9	1600	.0151	.05	.0323	.0224	.0488	.0322	.0224	.0168	.0348	.054	.0342	.0214	.03	.07
10	100	.3567	13.96	11.39	8.448	9.087	8.951	7.482	6.587	6.299	10.32	3.056	11.11	.01	.25
10	400	.0072	1.773	.7039	.1262	1.502	.6557	.1241	5.536	5.684	8.802	.0202	2.288	.01	.34
10	1600	.0082	.0132	.0053	.0032	.0132	.0053	.0032	.0116	.1161	.1171	.0081	.0058	.02	.06
11	100	.0377	.3824	.2333	.1324	.1205	.1188	.0979	.0528	.0463	.1263	.1362	.1922	.03	.14
11	400	.0214	.0611	.0337	.0219	.0446	.0299	.0211	.014	.0148	.0349	.0593	.0299	.05	.08
11	1600	.0252	.0197	.0098	.0067	.0182	.0096	.0067	.0044	.007	.017	.0596	.0067	.04	.08
12	100	.3658	3.51	2.787	2.195	2.136	2.108	1.903	1.525	1.457	2.6	.937	2.686	.01	.13
12	400	.0352	4.844	3.37	1.624	3.933	3.063	1.563	2.058	2.532	3.489	.122	3.528	.03	.09
12	1600	.0168	.6106	.5242	.3905	.6104	.5242	.3905	.7642	1.519	1.378	.0382	.4264	.01	.08
13	100	.0302	.4241	.2456	.14	.1475	.1352	.1069	.056	.0568	.1304	.1426	.2101	.03	.08
13	400	.09	.14	.087	.061	.1195	.0827	.0602	.0383	.054	.0965	.1241	.0796	.04	.06
13	1600	.0255	1.995	1.75	1.577	1.971	1.747	1.577	1.485	1.908	1.939	.168	1.528	.01	.23
14	100	.0449	1.235	.7393	.4249	1.222	.739	.4249	.6223	1.265	1.405	.1364	.9101	.02	.15
14	400	.0242	.8003	.6066	.398	.8002	.6066	.398	.914	2.099	1.984	.0559	.6105	.01	.06
14	1600	.0112	.867	.7479	.5654	.867	.7479	.5654	1.831	4.225	3.86	.0579	.7215	0	.06
15	100	.044	.0919	.0437	.0264	.0918	.0437	.0264	.1425	.4839	.2277	.0794	.0643	.03	.33
15	400	.0334	1.891	1.622	1.451	1.890	1.622	1.451	1.586	2.966	2.599	.1167	1.534	.01	.32
15	1600	.0041	.4359	.2451	.0517	.4359	.2451	.0517	2.014	3.551	3.109	.0135	.2793	0	.19

Table 2: Averages of $(ISE(\hat{h}) - MISE(h_0))/MISE(h_0)$ obtained for densities # 1 - 15 of Marron and Wand (1992). Sample sizes are $n = 100, 400$ and 1600 . v_{min} and v_{max} denote the minimum and maximum absolute value of the coefficient of variation for the 12 bandwidth selectors.

d	n	ISE	\hat{M}_0	\hat{M}_1	\hat{M}_2	\hat{N}_0	\hat{N}_1	\hat{N}_2	RN	SJ	PM	LSCV	FJ	v_{min}	v_{max}
1	100	-.0877	.1775	.1473	.1797	.099	.1136	.1602	.2775	.2707	.1899	.7877	.1346	.11	.39
1	400	-.077	.0737	.0637	.0692	.0513	.0552	.0641	.1064	.1104	.084	.4273	.0682	.13	.59
1	1600	-.0684	.0338	.0279	.0288	.0243	.0248	.0271	.044	.0463	.0394	.2259	.036	.16	1.08
2	100	-.0707	.1769	.145	.1755	.1066	.1187	.1636	.2469	.2345	.1683	.6116	.1372	.1	.43
2	400	-.0677	.0928	.0724	.0711	.0625	.0624	.067	.0822	.0804	.0846	.3061	.0684	.13	.48
2	1600	-.0658	.0451	.0361	.0361	.0344	.0333	.035	.0462	.0455	.0463	.2228	.041	.16	.76
3	100	-.0304	.5714	.4062	.2658	.5674	.4058	.2657	.2888	.4445	.6161	.1722	.4535	.04	.51
3	400	-.0269	.2352	.1712	.1212	.2352	.1712	.1212	.1997	.4615	.5829	.0776	.1746	.04	.6
3	1600	-.0179	.0926	.0662	.0521	.0926	.0662	.0521	.0945	.3299	.3919	.0522	.0618	.05	.77
4	100	-.0305	.5356	.3531	.2183	.4948	.3446	.2173	.1635	.2005	.3895	.1734	.4378	.08	.64
4	400	-.0282	.1212	.0839	.0676	.1208	.0839	.0676	.0627	.1018	.1782	.0906	.0883	.13	.62
4	1600	-.0332	.0572	.0419	.0374	.0572	.0419	.0374	.035	.0584	.0905	.0605	.0388	.24	.53
5	100	-.0675	.1673	.1296	.1293	.1031	.1069	.1193	.1754	.1824	.1489	.2312	.1215	.16	.45
5	400	-.068	.0861	.0655	.0651	.0568	.0557	.0607	.0789	.0823	.0826	.1439	.0626	.21	.5
5	1600	-.071	.0471	.0341	.0314	.0322	.0296	.0296	.031	.0326	.0452	.0986	.0301	.27	.85
6	100	-.0413	.4667	.3344	.2356	.2051	.2132	.1998	.134	.1098	.194	.4145	.2854	.05	.59
6	400	-.0362	.1581	.1115	.0922	.1259	.1036	.0904	.0784	.0705	.1096	.2711	.1054	.13	.64
6	1600	-.04	.0526	.0386	.0355	.0496	.0383	.0354	.0361	.0336	.0473	.1588	.0386	.17	.67
7	100	-.0346	.1135	.0773	.079	.1128	.0772	.079	.0572	.0782	.1052	.2895	.0767	.13	.69
7	400	-.0333	.0494	.037	.0377	.0493	.037	.0377	.0331	.0408	.0505	.1866	.0359	.15	.65
7	1600	-.0346	.0257	.0198	.0195	.0257	.0198	.0195	.0192	.0225	.0269	.1246	.0216	.2	1.07
8	100	-.0499	.4433	.3428	.2564	.2381	.2443	.2216	.12	.1024	.1935	.3687	.2946	.05	.43
8	400	-.0335	.273	.1968	.139	.2262	.1849	.1372	.095	.0976	.1641	.2076	.1664	.1	.63
8	1600	-.0308	.0796	.0577	.0504	.0777	.0575	.0504	.0452	.0496	.0777	.1342	.0538	.18	.63
9	100	-.0504	.4207	.2957	.2158	.2222	.2121	.1939	.1214	.1098	.1865	.3298	.2712	.05	.42
9	400	-.034	.2299	.1686	.1302	.2063	.1634	.1291	.0864	.1025	.1596	.1897	.1455	.11	.59
9	1600	-.0262	.1079	.0811	.0657	.1058	.0808	.0656	.053	.0784	.11	.1193	.0623	.17	.65
10	100	-.0254	.4634	.4356	.3668	.4264	.4173	.3596	.4245	.4224	.4421	.1803	.4469	.01	.53
10	400	-.0111	.4277	.2023	.0722	.4219	.2012	.0722	2.088	2.157	2.385	.0535	.6099	.01	1.27
10	1600	-.0143	.0403	.0241	.0194	.0403	.0241	.0194	.0374	.2186	.2208	.0331	.026	.08	.92
11	100	-.0405	.387	.2739	.1876	.1672	.174	.1597	.1135	.0952	.1587	.3432	.2332	.05	.5
11	400	-.0267	.0986	.0693	.058	.0795	.0648	.057	.0511	.0468	.0694	.1758	.067	.13	.59
11	1600	-.0186	.0251	.0177	.0155	.0237	.0175	.0155	.0147	.0155	.0228	.0919	.0168	.13	.58
12	100	-.048	.2685	.2435	.2121	.2282	.2231	.2032	.2123	.2096	.2477	.1539	.2473	.03	.24
12	400	-.0218	.7007	.5366	.3056	.6544	.5207	.3026	.4851	.58	.6605	.0901	.6122	.03	.48
12	1600	-.012	.2808	.251	.1999	.2807	.251	.1999	.3342	.5533	.5154	.0382	.2136	.02	.72
13	100	-.0262	.3056	.1979	.1419	.1378	.1304	.1232	.076	.0689	.1136	.2667	.1763	.05	.58
13	400	-.0474	.0761	.0549	.0469	.0665	.0528	.0465	.0355	.0391	.0567	.0908	.0523	.14	.31
13	1600	-.011	.4307	.4056	.3784	.4285	.4052	.3784	.3754	.4155	.4249	.0722	.3812	.02	.91
14	100	-.0217	.3116	.2186	.1468	.3103	.2185	.1468	.2003	.3305	.3535	.0963	.2556	.03	.48
14	400	-.0136	.2589	.2117	.1547	.2589	.2117	.1547	.2887	.5166	.4963	.042	.213	.02	.57
14	1600	-.0081	.3041	.2752	.227	.3041	.2752	.227	.5073	.8851	.8337	.0409	.2687	.01	.83
15	100	-.0207	.0704	.0424	.0335	.0704	.0424	.0335	.0661	.2242	.145	.1007	.0547	.06	.47
15	400	-.0104	.3662	.3445	.3279	.3662	.3445	.3279	.3407	.4417	.4168	.0544	.3369	.02	.91
15	1600	-.0053	.3273	.1927	.0607	.3273	.1927	.0607	1.272	1.684	1.587	.028	.2702	0	1.57

Table 3: Averages of $(MISE(\hat{h}) - MISE(h_0))/MISE(h_0)$ for densities # 1 - 15 of Marron and Wand (1992). Sample sizes are $n = 100, 400$ and 1600 . v_{min} and v_{max} denote the minimum and maximum value of the coefficient of variation obtained for the 12 bandwidth selectors.

d	n	ISE	\hat{M}_0	\hat{M}_1	\hat{M}_2	\hat{N}_0	\hat{N}_1	\hat{N}_2	RN	SJ	PM	LSCV	FJ	v_{min}	v_{max}
1	100	.0836	.0866	.0482	.0505	.0217	.0252	.0401	.0928	.0973	.0619	.3975	.0393	.04	.17
1	400	.0726	.0245	.0123	.0114	.0066	.0074	.0094	.0247	.0282	.0184	.2213	.0104	.05	.13
1	1600	.0644	.0081	.0033	.0029	.0019	.002	.0024	.0072	.0081	.0061	.1166	.0042	.04	.13
2	100	.0699	.1109	.0632	.0596	.0329	.035	.0486	.0845	.0833	.0681	.328	.0552	.05	.11
2	400	.0656	.0369	.0189	.0152	.014	.0126	.013	.0186	.0188	.0236	.1593	.0138	.04	.1
2	1600	.0616	.0123	.0056	.0046	.0053	.0042	.0042	.0069	.0067	.0087	.1023	.005	.04	.11
3	100	.0319	.5378	.3684	.2296	.5336	.368	.2296	.2513	.4111	.5859	.1294	.4164	.02	.07
3	400	.0217	.2126	.1464	.0974	.2126	.1464	.0974	.1766	.4427	.5655	.0581	.151	.02	.06
3	1600	.0205	.0659	.0408	.0281	.0659	.0408	.0281	.0688	.3031	.3644	.0286	.0368	.01	.06
4	100	.0285	.5174	.3204	.174	.4696	.3098	.1726	.1116	.1556	.3609	.1144	.4076	.05	.09
4	400	.0275	.0818	.0461	.031	.0814	.0461	.031	.0259	.0622	.1351	.0453	.0492	.03	.07
4	1600	.0331	.0244	.0125	.0092	.0244	.0125	.0092	.0078	.0266	.0548	.0236	.0097	.03	.06
5	100	.0658	.0882	.0472	.038	.0312	.0299	.032	.0593	.0686	.0543	.1052	.0413	.05	.08
5	400	.0713	.0295	.0145	.0121	.0107	.0096	.0105	.0189	.0225	.0209	.0569	.011	.04	.09
5	1600	.0763	.0102	.0042	.0033	.0037	.003	.003	.0046	.006	.0078	.0389	.0032	.04	.09
6	100	.0503	.4313	.2882	.1803	.1689	.1662	.1415	.077	.0638	.1629	.2693	.2467	.02	.08
6	400	.0364	.1149	.0658	.0433	.0843	.0585	.0418	.0274	.0268	.0659	.1466	.0576	.05	.1
6	1600	.0394	.0232	.0104	.0073	.021	.0102	.0073	.0061	.0067	.0193	.0846	.0077	.05	.11
7	100	.0362	.0727	.0346	.0269	.0721	.0345	.0269	.0161	.0467	.0699	.1557	.0344	.04	.11
7	400	.033	.0232	.0106	.0085	.0232	.0106	.0085	.006	.0211	.0286	.0935	.0078	.03	.1
7	1600	.0346	.0084	.0034	.0026	.0084	.0034	.0026	.0023	.0091	.0122	.0648	.0028	.03	.11
8	100	.0557	.4269	.3133	.2194	.2175	.2122	.1815	.0891	.0791	.1909	.2683	.2763	.02	.09
8	400	.0319	.2205	.142	.0848	.1768	.131	.0832	.0463	.0572	.1233	.1162	.1153	.04	.1
8	1600	.0293	.0491	.0268	.0185	.0474	.0266	.0185	.0141	.0241	.0517	.0694	.0209	.04	.11
9	100	.0504	.411	.2704	.1721	.1983	.1785	.1454	.0756	.0742	.1663	.2221	.2484	.03	.07
9	400	.034	.1865	.1236	.0844	.1649	.1189	.0835	.0493	.0712	.1254	.1147	.1045	.03	.08
9	1600	.0262	.0778	.0509	.0356	.0761	.0507	.0356	.027	.0552	.0842	.0675	.0342	.03	.09
10	100	.0336	.4546	.4251	.3611	.4204	.4087	.3548	.4108	.4103	.4339	.1646	.4381	0	.11
10	400	.012	.4419	.2035	.0595	.4356	.2024	.0594	2.046	2.121	2.35	.0307	.63	0	.2
10	1600	.0149	.0229	.0093	.0057	.0229	.0093	.0057	.0202	.1938	.1954	.0144	.0101	.02	.06
11	100	.0481	.3489	.2296	.1406	.1331	.1305	.11	.0649	.0555	.1314	.2162	.1963	.02	.08
11	400	.0281	.0692	.0388	.0257	.0514	.0347	.0248	.017	.0176	.0403	.0951	.0347	.05	.1
11	1600	.0177	.0131	.0065	.0045	.012	.0064	.0044	.0029	.0046	.0113	.0546	.0045	.04	.1
12	100	.0475	.2415	.2175	.1903	.2032	.1986	.182	.1715	.1715	.2149	.1311	.2188	0	.07
12	400	.0222	.7186	.5517	.3106	.6686	.5338	.3069	.4807	.5835	.6694	.0632	.6187	.02	.07
12	1600	.0127	.2692	.2382	.186	.2691	.2381	.186	.3251	.5468	.5083	.0244	.2022	.01	.07
13	100	.0349	.2928	.1816	.1132	.1226	.111	.0917	.0492	.049	.1023	.1853	.1622	.03	.09
13	400	.0361	.058	.0355	.025	.0492	.0337	.0246	.0152	.0216	.0394	.069	.0324	.04	.07
13	1600	.0141	.4149	.3909	.3663	.4128	.3906	.3663	.361	.4011	.41	.0686	.3662	0	.11
14	100	.0218	.2948	.2002	.1272	.2935	.2001	.1272	.1815	.3134	.3396	.0682	.2377	.01	.08
14	400	.0141	.2422	.1946	.1368	.2422	.1946	.1368	.2724	.4969	.4783	.0282	.197	.01	.06
14	1600	.009	.2841	.255	.2072	.2841	.255	.2072	.4869	.8624	.8116	.0303	.2493	0	.06
15	100	.0299	.058	.0281	.0171	.0579	.0281	.0171	.0535	.2187	.1379	.0601	.0408	.02	.14
15	400	.0141	.3687	.3463	.329	.3687	.3463	.329	.3437	.4482	.4221	.0479	.3384	0	.13
15	1600	.0059	.32	.1842	.0496	.32	.1842	.0496	1.264	1.681	1.583	.0173	.2563	0	.12

Table 4: Averages of $[(\hat{h}-h_0)/h_0]^2$ obtained for densities # 1 - 15 of Marron and Wand (1992). Sample sizes are $n = 100, 400$ and 1600 . v_{min} and v_{max} denote the minimum and maximum value of the coefficient of variation for the 12 bandwidth selectors (N_2^κ based on the normal mixture reference $\hat{\eta} = \hat{\omega}\phi_{\hat{\mu}_1, \hat{\sigma}_1} + (1 - \hat{\omega})\phi_{\hat{\mu}_2, \hat{\sigma}_2}$, parameters estimated by maximum likelihood).

d	n	ISE	\hat{M}_0	\hat{M}_1	\hat{M}_2	$\hat{N}_{0,B}$	$\hat{N}_{1,B}$	$\hat{N}_{2,B}$	RN	SJ	PM	LSCV	FJ	v_{min}	v_{max}
1	100	.0322	.0373	.02	.0185	.0285	.0277	.0281	.0313	.0332	.0256	.0975	.0164	.04	.11
1	400	.0293	.0103	.0051	.0047	.0122	.0121	.0122	.0096	.011	.0077	.0608	.0043	.05	.15
1	1600	.0268	.0035	.0014	.0012	.003	.003	.0031	.003	.0034	.0026	.0387	.0018	.03	.17
2	100	.0299	.0536	.0296	.0253	.0218	.0215	.0239	.0316	.0315	.0317	.0966	.0257	.05	.1
2	400	.0279	.0166	.0085	.0067	.0041	.0044	.005	.0078	.0079	.0106	.0541	.0061	.04	.15
2	1600	.0266	.0055	.0025	.002	.0012	.0013	.0016	.003	.0029	.0039	.0371	.0022	.04	.1
3	100	.0333	1.3	.7883	.4339	.3849	.3588	.2792	.4412	.8098	1.333	.1617	.901	.03	.1
3	400	.015	.2109	.1399	.0896	.1503	.1175	.0827	.167	.4735	.6374	.0487	.1434	.02	.07
3	1600	.0122	.0443	.027	.0184	.0415	.0265	.0183	.046	.2184	.2664	.0186	.0243	.01	.06
4	100	.0208	1.067	.5996	.2712	.2103	.1659	.1009	.132	.183	.5251	.1094	.7498	.08	.31
4	400	.0159	.0528	.0292	.0193	.0097	.0099	.0102	.016	.0396	.0888	.0266	.0312	.04	.08
4	1600	.0173	.0134	.0068	.005	.0029	.003	.0031	.0042	.0146	.0303	.0127	.0053	.03	.07
5	100	.0274	.0412	.0215	.0168	.0081	.0092	.0112	.0229	.0262	.0247	.0441	.0188	.05	.07
5	400	.0296	.0131	.0064	.0053	.0024	.0029	.0037	.0078	.0093	.0092	.0245	.0048	.05	.09
5	1600	.0322	.0045	.0019	.0015	.0008	.001	.0011	.002	.0026	.0035	.0172	.0014	.04	.12
6	100	.0318	.3966	.2462	.1426	.0234	.0245	.0286	.0517	.044	.1307	.1366	.2027	.03	.12
6	400	.0179	.0654	.0368	.0237	.007	.0077	.009	.0144	.0144	.0367	.0587	.0319	.05	.07
6	1600	.0189	.0116	.0052	.0036	.0018	.002	.0022	.003	.0033	.0096	.0334	.0038	.04	.09
7	100	.0177	.0399	.0185	.0138	.006	.0065	.0085	.0083	.0253	.0383	.0591	.0184	.04	.08
7	400	.0157	.0116	.0053	.0041	.0018	.0021	.0028	.0029	.0105	.0142	.0373	.0038	.04	.08
7	1600	.0163	.004	.0016	.0013	.0007	.0008	.0009	.0011	.0043	.0059	.0265	.0013	0	.12
8	100	.0578	.6185	.4175	.2707	.0575	.0597	.0637	.0916	.0828	.2356	.1912	.3531	.02	.15
8	400	.0184	.1673	.1039	.0594	.0172	.0178	.0188	.0308	.0387	.0877	.0613	.082	.04	.07
8	1600	.0154	.0276	.0149	.0102	.0049	.0051	.0054	.0077	.0134	.029	.031	.0115	.04	.08
9	100	.0399	.5395	.3318	.1943	.0214	.024	.0314	.0774	.0766	.1906	.1636	.2941	.03	.09
9	400	.0223	.1608	.103	.0685	.0271	.0302	.0321	.0387	.0563	.1036	.0666	.086	.04	.07
9	1600	.0151	.05	.0323	.0224	.0243	.0226	.019	.0168	.0348	.054	.0342	.0214	.03	.07
10	100	.3567	13.96	11.39	8.448	6.295	6.234	5.4	6.587	6.299	10.32	3.056	11.11	.01	.25
10	400	.0072	1.773	.7039	.1262	1.096	.4019	.0838	5.536	5.684	8.802	.0202	2.288	.01	.39
10	1600	.0082	.0132	.0053	.0032	.0117	.0048	.003	.0116	.1161	.1171	.0081	.0058	.02	.07
11	100	.0377	.3824	.2333	.1324	.025	.0258	.029	.0528	.0463	.1263	.1362	.1922	.03	.14
11	400	.0214	.0611	.0337	.0219	.0071	.0076	.0088	.014	.0148	.0349	.0593	.0299	.05	.08
11	1600	.0252	.0197	.0098	.0067	.003	.0033	.0036	.0044	.007	.017	.0596	.0067	.04	.08
12	100	.3658	3.51	2.787	2.195	.9649	.9809	.9783	1.525	1.457	2.6	.9372	2.686	.01	.13
12	400	.0352	4.844	3.37	1.624	2.146	1.8	1.076	2.058	2.532	3.489	.1217	3.528	.03	.09
12	1600	.0168	.6106	.5242	.3905	.601	.5214	.3903	.7642	1.519	1.3781	.0382	.4264	.01	.08
13	100	.0302	.4241	.2456	.14	.0229	.0239	.029	.056	.0568	.1304	.1426	.2101	.03	.08
13	400	.09	.14	.087	.061	.0257	.0273	.03	.0383	.054	.0965	.1241	.0796	.04	.06
13	1600	.0255	1.995	1.75	1.577	1.515	1.523	1.467	1.485	1.908	1.939	.1676	1.528	.01	.23
14	100	.0449	1.235	.7393	.4249	.6464	.5305	.3682	.6223	1.265	1.405	.1364	.9101	.02	.15
14	400	.0242	.8003	.6066	.398	.7661	.5991	.3967	.9142	2.099	1.984	.0559	.6105	.01	.06
14	1600	.0112	.867	.7479	.5654	.8659	.7476	.5654	1.831	4.225	3.86	.0579	.7215	0	.06
15	100	.044	.0919	.0437	.0264	.0728	.0366	.0232	.1425	.4839	.2277	.0794	.0643	.03	.33
15	400	.0334	1.891	1.622	1.451	1.826	1.588	1.432	1.586	2.966	2.599	.1167	1.534	.01	.32
15	1600	.0041	.4359	.2451	.0517	.4312	.2444	.0516	2.014	3.551	3.109	.0135	.2793	0	.19

Table 5: Averages of $(ISE(\hat{h}) - MISE(h_0))/MISE(h_0)$ obtained for densities # 1 - 15 of Marron and Wand (1992). Sample sizes are $n = 100, 400$ and 1600 . v_{min} and v_{max} denote the minimum and maximum absolute value of the coefficient of variation for the 12 bandwidth selectors (N_2^κ based on the normal mixture reference $\hat{\eta} = \hat{\omega}\phi_{\hat{\mu}_1, \hat{\sigma}_1} + (1 - \hat{\omega})\phi_{\hat{\mu}_2, \hat{\sigma}_2}$, parameters estimated by maximum likelihood).

d	n	ISE	\hat{M}_0	\hat{M}_1	\hat{M}_2	$\hat{N}_{0,B}$	$\hat{N}_{1,B}$	$\hat{N}_{2,B}$	RN	SJ	PM	LSCV	FJ	v_{min}	v_{max}
1	100	-.0877	.1775	.1473	.1797	.2547	.2564	.2709	.2775	.2707	.1899	.7877	.1346	.11	.38
1	400	-.077	.0737	.0637	.0692	.1156	.1171	.1205	.1064	.1104	.084	.4273	.0682	.13	.49
1	1600	-.0684	.0338	.0279	.0288	.0389	.0387	.0402	.044	.0463	.0394	.2259	.036	.16	.95
2	100	-.0707	.1769	.145	.1755	.1625	.1655	.1952	.2469	.2345	.1683	.6116	.1372	.1	.43
2	400	-.0677	.0928	.0724	.0711	.0548	.0582	.0645	.0822	.0804	.0846	.3061	.0684	.13	.55
2	1600	-.0658	.0451	.0361	.0361	.0287	.03	.0327	.0462	.0455	.0463	.2228	.041	.16	.87
3	100	-.0304	.5714	.4062	.2658	.2613	.2483	.2074	.2888	.4445	.6161	.1722	.4535	.04	.51
3	400	-.0269	.2352	.1712	.1212	.1832	.1514	.1152	.1997	.4615	.5829	.0776	.1746	.04	.6
3	1600	-.0179	.0926	.0662	.0521	.0885	.0654	.0519	.0945	.3299	.3919	.0522	.0618	.05	.77
4	100	-.0305	.5356	.3531	.2183	.1888	.1731	.1486	.1635	.2005	.3895	.1734	.4378	.08	.64
4	400	-.0282	.1212	.0839	.0676	.0487	.049	.0504	.0627	.1018	.1782	.0906	.0883	.13	.62
4	1600	-.0332	.0572	.0419	.0374	.029	.0295	.0306	.035	.0584	.0905	.0605	.0388	.24	.63
5	100	-.0675	.1673	.1296	.1293	.0805	.0908	.1082	.1754	.1824	.1489	.2312	.1215	.16	.45
5	400	-.068	.0861	.0655	.0651	.042	.0463	.0547	.0789	.0823	.0826	.1439	.0626	.21	.65
5	1600	-.071	.0471	.0341	.0314	.0236	.0251	.027	.031	.0326	.0452	.0986	.0301	.27	1.05
6	100	-.0413	.4667	.3344	.2356	.0845	.0891	.107	.134	.1098	.194	.4145	.2854	.05	.59
6	400	-.0362	.1581	.1115	.0922	.049	.0525	.0593	.0784	.0705	.1096	.2711	.1054	.13	.64
6	1600	-.04	.0526	.0386	.0355	.0264	.0279	.0298	.0361	.0336	.0473	.1588	.0386	.17	.85
7	100	-.0346	.1135	.0773	.079	.046	.0507	.0652	.0572	.0782	.1052	.2895	.0767	.13	.69
7	400	-.0333	.0494	.037	.0377	.025	.0274	.0322	.0331	.0408	.0505	.1866	.0359	.15	.85
7	1600	-.0346	.0257	.0198	.0195	.015	.0159	.0174	.0192	.0225	.0269	.1246	.0216	.2	1.37
8	100	-.0499	.4433	.3428	.2564	.1156	.1197	.1285	.12	.1024	.1935	.3687	.2946	.05	.43
8	400	-.0335	.273	.1968	.139	.0734	.0757	.08	.095	.0976	.1641	.2076	.1664	.1	.63
8	1600	-.0308	.0796	.0577	.0504	.038	.0388	.0406	.0452	.0496	.0777	.1342	.0538	.18	.63
9	100	-.0504	.4207	.2957	.2158	.0694	.0759	.0966	.1214	.1098	.1865	.3298	.2712	.05	.42
9	400	-.034	.2299	.1686	.1302	.0673	.0744	.0811	.0864	.1025	.1596	.1897	.1455	.11	.59
9	1600	-.0262	.1079	.0811	.0657	.0634	.063	.0591	.053	.0784	.11	.1193	.0623	.17	.65
10	100	-.0254	.4634	.4356	.3668	.3917	.3831	.3277	.4245	.4224	.4421	.1803	.4469	.01	.53
10	400	-.0111	.4277	.2023	.0722	.3447	.1483	.0603	2.088	2.157	2.385	.0535	.6099	.01	1.27
10	1600	-.0143	.0403	.0241	.0194	.0369	.0226	.0186	.0374	.2186	.2208	.0331	.026	.08	.92
11	100	-.0405	.387	.2739	.1876	.0758	.0787	.0887	.1135	.0952	.1587	.3432	.2332	.05	.5
11	400	-.0267	.0986	.0693	.058	.0325	.0344	.0388	.0511	.0468	.0694	.1758	.067	.13	.59
11	1600	-.0186	.0251	.0177	.0155	.0109	.0114	.0121	.0147	.0155	.0228	.0919	.0168	.13	.77
12	100	-.048	.2685	.2435	.2121	.1632	.1611	.1538	.2123	.2096	.2477	.1539	.2473	.03	.24
12	400	-.0218	.7007	.5366	.3056	.4707	.3974	.2586	.4851	.58	.6605	.0901	.6122	.03	.48
12	1600	-.012	.2808	.251	.1999	.2776	.2501	.1998	.3342	.5533	.5154	.0382	.2136	.02	.72
13	100	-.0262	.3056	.1979	.1419	.0506	.0527	.0705	.076	.0689	.1136	.2667	.1763	.05	.58
13	400	-.0474	.0761	.0549	.0469	.0251	.0267	.0313	.0355	.0391	.0567	.0908	.0523	.14	.42
13	1600	-.011	.4307	.4056	.3784	.3797	.3801	.3658	.3754	.4155	.4249	.0722	.3812	.02	.91
14	100	-.0217	.3116	.2186	.1468	.2045	.1763	.1347	.2003	.3305	.3535	.0963	.2556	.03	.48
14	400	-.0136	.2589	.2117	.1547	.2512	.2099	.1544	.2887	.5166	.4963	.042	.213	.02	.57
14	1600	-.0081	.3041	.2752	.227	.3038	.2751	.227	.5073	.8851	.8337	.0409	.2687	.01	.83
15	100	-.0207	.0704	.0424	.0335	.0593	.0381	.0315	.0661	.2242	.145	.1007	.0547	.06	.47
15	400	-.0104	.3662	.3445	.3279	.3613	.3417	.3263	.3407	.4417	.4168	.0544	.3369	.02	.91
15	1600	-.0053	.3273	.1927	.0607	.3254	.1925	.0607	1.2716	1.6836	1.587	.028	.2702	0	1.57

Table 6: Averages of $(MISE(\hat{h}) - MISE(h_0))/MISE(h_0)$ for densities # 1 - 15 of Marron and Wand (1992). Sample sizes are $n = 100, 400$ and 1600 . v_{min} and v_{max} denote the minimum and maximum value of the coefficient of variation obtained for the 12 bandwidth selectors (N_2^K based on the normal mixture reference $\hat{\eta} = \hat{\omega}\phi_{\hat{\mu}_1, \hat{\sigma}_1} + (1 - \hat{\omega})\phi_{\hat{\mu}_2, \hat{\sigma}_2}$, parameters estimated by maximum likelihood).

d	n	ISE	\hat{M}_0	\hat{M}_1	\hat{M}_2	$\hat{N}_{0,B}$	$\hat{N}_{1,B}$	$\hat{N}_{2,B}$	RN	SJ	PM	LSCV	FJ	v_{min}	v_{max}
1	100	.0836	.0866	.0482	.0505	.0951	.0931	.0958	.0928	.0973	.0619	.3975	.0393	.04	.15
1	400	.0726	.0245	.0123	.0114	.0355	.0353	.0354	.0247	.0282	.0184	.2213	.0104	.05	.18
1	1600	.0644	.0081	.0033	.0029	.0075	.0075	.0076	.0072	.0081	.0061	.1166	.0042	.04	.18
2	100	.0699	.1109	.0632	.0596	.0594	.0582	.0658	.0845	.0833	.0681	.328	.0552	.05	.12
2	400	.0656	.0369	.0189	.0152	.0099	.0106	.0121	.0186	.0188	.0236	.1593	.0138	.04	.19
2	1600	.0616	.0123	.0056	.0046	.0027	.003	.0035	.0069	.0067	.0087	.1023	.005	.04	.11
3	100	.0319	.5378	.3684	.2296	.2136	.2016	.1649	.2513	.4111	.5859	.1294	.4164	.02	.07
3	400	.0217	.2126	.1464	.0974	.1576	.1254	.0907	.1766	.4427	.5655	.0581	.151	.02	.06
3	1600	.0205	.0659	.0408	.0281	.062	.04	.0279	.0688	.3031	.3644	.0286	.0368	.01	.06
4	100	.0285	.5174	.3204	.174	.128	.1128	.0892	.1116	.1556	.3609	.1144	.4076	.05	.13
4	400	.0275	.0818	.0461	.031	.0167	.0169	.0174	.0259	.0622	.1351	.0453	.0492	.03	.07
4	1600	.0331	.0244	.0125	.0092	.0053	.0055	.0059	.0078	.0266	.0548	.0236	.0097	.03	.07
5	100	.0658	.0882	.0472	.038	.0189	.0215	.0264	.0593	.0686	.0543	.1052	.0413	.05	.08
5	400	.0713	.0295	.0145	.0121	.0056	.0066	.0086	.0189	.0225	.0209	.0569	.011	.04	.09
5	1600	.0763	.0102	.0042	.0033	.0019	.0022	.0025	.0046	.006	.0078	.0389	.0032	.04	.09
6	100	.0503	.4313	.2882	.1803	.0416	.0428	.0496	.077	.0638	.1629	.2693	.2467	.02	.08
6	400	.0364	.1149	.0658	.0433	.0137	.0151	.0176	.0274	.0268	.0659	.1466	.0576	.05	.1
6	1600	.0394	.0232	.0104	.0073	.0037	.0041	.0046	.0061	.0067	.0193	.0846	.0077	.05	.11
7	100	.0362	.0727	.0346	.0269	.0121	.0133	.0176	.0161	.0467	.0699	.1557	.0344	.04	.11
7	400	.033	.0232	.0106	.0085	.0038	.0045	.0057	.006	.0211	.0286	.0935	.0078	.03	.1
7	1600	.0346	.0084	.0034	.0026	.0014	.0016	.0019	.0023	.0091	.0122	.0648	.0028	.03	.11
8	100	.0557	.4269	.3133	.2194	.0873	.0884	.0925	.0891	.0791	.1909	.2683	.2763	.02	.09
8	400	.0319	.2205	.142	.0848	.0298	.0306	.0322	.0463	.0572	.1233	.1162	.1153	.04	.1
8	1600	.0293	.0491	.0268	.0185	.0094	.0097	.0103	.0141	.0241	.0517	.0694	.0209	.04	.11
9	100	.0504	.411	.2704	.1721	.0276	.0305	.0406	.0756	.0742	.1663	.2221	.2484	.03	.07
9	400	.034	.1865	.1236	.0844	.0358	.0398	.0423	.0493	.0712	.1254	.1147	.1045	.03	.08
9	1600	.0262	.0778	.0509	.0356	.0388	.036	.0305	.027	.0552	.0842	.0675	.0342	.03	.09
10	100	.0336	.4546	.4251	.3611	.3833	.3734	.3239	.4108	.4103	.4339	.1646	.4381	0	.11
10	400	.012	.4419	.2035	.0595	.3511	.1433	.0458	2.046	2.121	2.35	.0307	.63	0	.21
10	1600	.0149	.0229	.0093	.0057	.0203	.0084	.0053	.0202	.1938	.1954	.0144	.0101	.02	.06
11	100	.0481	.3489	.2296	.1406	.038	.0383	.0422	.0649	.0555	.1314	.2162	.1963	.02	.08
11	400	.0281	.0692	.0388	.0257	.0088	.0094	.0109	.017	.0176	.0403	.0951	.0347	.05	.1
11	1600	.0177	.0131	.0065	.0045	.002	.0022	.0024	.0029	.0046	.0113	.0546	.0045	.04	.1
12	100	.0475	.2415	.2175	.1903	.1404	.139	.1336	.1715	.1715	.2149	.1311	.2188	0	.07
12	400	.0222	.7186	.5517	.3106	.4686	.3964	.254	.4807	.5835	.6694	.0632	.6187	.02	.07
12	1600	.0127	.2692	.2382	.186	.266	.2372	.186	.3251	.5468	.5083	.0244	.2022	.01	.07
13	100	.0349	.2928	.1816	.1132	.0249	.0255	.0331	.0492	.049	.1023	.1853	.1622	.03	.13
13	400	.0361	.058	.0355	.025	.01	.0107	.0121	.0152	.0216	.0394	.069	.0324	.04	.08
13	1600	.0141	.4149	.3909	.3663	.3665	.367	.3545	.361	.4011	.41	.0686	.3662	0	.11
14	100	.0218	.2948	.2002	.1272	.1873	.1576	.1146	.1815	.3134	.3396	.0682	.2377	.01	.08
14	400	.0141	.2422	.1946	.1368	.2348	.1928	.1365	.2724	.4969	.4783	.0282	.197	.01	.06
14	1600	.009	.2841	.255	.2072	.2839	.255	.2072	.4869	.8624	.8116	.0303	.2493	0	.06
15	100	.0299	.058	.0281	.0171	.0461	.0236	.0151	.0535	.2187	.1379	.0601	.0408	.02	.14
15	400	.0141	.3687	.3463	.329	.3632	.3432	.3272	.3437	.4482	.4221	.0479	.3384	0	.13
15	1600	.0059	.32	.1842	.0496	.318	.184	.0496	1.264	1.6812	1.5832	.0173	.2563	0	.12